

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Integrative analysis of the metabolic signatures of ageing and age-related diseases

Zierer, Jonas

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Integrative analysis of the metabolic signatures of ageing and age-related diseases

Jonas Zierer



King's College London

Department of Twin Research and Genetic Epidemiology
2017

Abstract

Ageing is a complex process and is the strongest risk factor for many diseases. To elucidate processes underlying biological ageing, many studies have investigated the associations of individual metabolites or genes with age and age-related phenotypes. While these studies improved our understanding of the pathogenic mechanisms for many diseases, they have their limitations. Analysing phenotypes separately in classical association studies typically neglects relationships and interactions between and across different phenotypes (e.g. comorbidities), genes (e.g. pleiotropy), and metabolites (e.g. through shared biochemical pathways). The aim of this thesis was to better understand age-related phenotypes and their interdependencies through exploring shared underlying processes. To this end, I first identified biomarkers of ageing and age-related diseases, focusing on chronic kidney disease, using metabolomics and glycomics technologies. I identified several metabolites associated with leukocyte telomere length, a common marker of biological ageing. Then I investigated the associations of molecular phenotypes with renal disease. Analysing metabolomic profiles associated with renal function in diabetic and non-diabetic cohorts illustrated similarities between the different aetiologies of kidney disease, such as the lack of renal conversion of amino acids, but also differences, particularly of lipid and energy metabolism. Subsequent analyses identified changes of Immunoglobulin G glycosylation as a novel inflammatory pathway involved in renal disease. Then, I assessed the potential of the faecal metabolome as a functional readout of the gut microbial community to investigate its association with biological ageing. While faecal metabolites were only moderately associated with age and renal function, they showed great potential as novel profiling method for studying the microbiome, particularly with respect to obesity. Next, I integrated metabolomics data from plasma, urine, and saliva to model cross-fluid metabolism individually for kidney disease patients and healthy controls. By comparing both models, I identified metabolic key processes impaired in kidney disease. Finally, I integrated metabolomic and glycomic biomarkers of ageing with other omics markers as well as extensive phenotypic data to investigate their multivariate interdependencies, underlying the comorbidities of age-related diseases. This comprehensive integration of age-related phenotypes highlighted several molecular mechanisms that potentially cause the joint occurrence of diseases with age. Considering the complex aetiologies of different diseases and their dependencies will be needed to facilitate personalised healthcare. In conclusion, I have shown the future potential of systems and network biology approaches for understanding disease mechanisms and precision medicine.

Acknowledgements

First of all, I want to express my heartfelt gratitude to my three supervisors, Cristina Menni, Gabi Kastenmüller, and Tim Spector. I consider myself very lucky to have had three exceptionally supportive advisers with so diverse backgrounds and skill sets. I particularly want to thank Gabi for offering me this opportunity and paving my way into science in the first place; Cristina for keeping me on track to publish my research by setting these overly optimistic deadlines; and Tim for the entire TwinsUK experience.

Besides my supervisors, I am indebted to many colleagues and collaborators, who helped me with their support and expertise. First of all, I want to kindly thank Clara Barrios for teaching me the basics of nephrology. Moreover, I want to thank Jan Krumsiek for all the discussions on graphical models, Matt Jackson for helping me getting started in the microbiome world, Massimo Mangino for the advice about GWASs, and Rob Mohnney, Peter Würtz, and Gordon Lauc for their patience in explaining the experimental details of metabolomics and glycomics, respectively, to me. I also want to thank all the other members of TwinsUK for the continuous, joint effort collecting, cleaning, and normalising all this data, as well as the volunteers participating in the TwinsUK and KORA studies without whom none of this research would be possible.

Publications

- **Zierer, J.**, Menni, C., Kastenmüller, G. and Spector, T. D. (2015). 'Integration of 'omics' data in aging research: from biomarkers to systems biology.' *Aging cell* 14.6, pp. 933–44
- **Zierer, J.**, Kastenmüller, G., Suhre, K., Gieger, C., Codd, V., Tsai, P.-C., Bell, J., Peters, A., Strauch, K., Schulz, H., Weidinger, S., Mohny, R. P., Samani, N. J., Spector, T., Mangino, M. and Menni, C. (2016a). 'Metabolomics profiling reveals novel markers for leukocyte telomere length.' *Aging* 8.1, pp. 77–94
- Barrios*, C., **Zierer***, J., Gudelj, I., Štambuk, J., Ugrina, I., Rodríguez, E., Soler, M. J., Pavić, T., Šimurina, M., Keser, T., Pučić-Baković, M., Mangino, M., Pascual, J., Spector, T. D., Lauc, G. and Menni, C. (2016). 'Glycosylation Profile of IgG in Moderate Kidney Dysfunction.' *Journal of the American Society of Nephrology : JASN* 27.3, pp. 933–41
- **Zierer, J.**, Pallister, T., Tsai, P.-C., Krumsiek, J., Bell, J. T., Lauc, G., Spector, T. D., Menni, C. and Kastenmüller, G. (2016b). 'Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model'. *Scientific Reports* 6.1, p. 37646
- Menni, C., **Zierer, J.**, Valdes, A. M. and Spector, T. D. (2017a). 'Mixing omics: combining genetics and metabolomics to study rheumatic diseases.' *Nature reviews. Rheumatology* 13.3, pp. 174–181
- Keser, T., Vučković, F., Barrios, C., **Zierer, J.**, Wahl, A., Akinkuolie, A. O., Štambuk, J., Nakić, N., Pavić, T., Periša, J., Mora, S., Gieger, C., Menni, C., Spector, T. D., Gornik, O. and Lauc, G. (2017a). 'Effects of statins on the immunoglobulin G glycome'. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1861.5, pp. 1152–1158
- Long, T., Hicks, M., Yu, H.-C., Biggs, W. H., Kirkness, E. F., Menni, C., **Zierer, J.**, Small, K. S., Mangino, M., Messier, H., Brewerton, S., Turpaz, Y., Perkins, B. A., Evans, A. M., Miller, L. A. D., Guo, L., Caskey, C. T., Schork, N. J., Garner, C., Spector, T. D., Venter, J. C. and Telenti, A. (2017b). 'Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites'. *Nature Genetics* 49.4, pp. 568–578

Contents

Abstract	2
Acknowledgements	3
Publications	4
List of Figures	10
List of Tables	12
Glossary	14
1 Introduction	18
1.1 Ageing	19
1.1.1 Theories of ageing	20
1.1.2 Biological ageing	21
1.1.3 Chronic kidney disease	23
1.2 ‘Omics’	27
1.2.1 Genomics	27
1.2.2 Epigenomics	27
1.2.3 Transcriptomics	29
1.2.4 Proteomics	30
1.2.5 Post-translational modifications - glycomics	31
1.2.6 Metabolomics	31
1.2.7 Microbiomics	32
1.2.8 Phenomics	33
1.3 Systems biology	33
1.3.1 Introduction to graphs	35
1.3.2 Analysis of pre-defined graphs	36
1.3.3 Graph inference	38
1.3.4 Model biological systems	41
1.4 Conclusion	41
2 Hypothesis, aims, and outline	43
2.1 Hypothesis	43
2.2 Aims	43
2.3 Outline	44

3	Material and methods	45
3.1	Study populations	45
3.1.1	TwinsUK	45
3.1.2	KORA	48
3.1.3	SHIP and GANI_MED	49
3.1.4	YoungFinns	49
3.2	Phenotypic measures	49
3.2.1	Anthropometry	49
3.2.2	Body composition	49
3.2.3	Renal function	50
3.2.4	Leukocyte telomere length	51
3.2.5	Liver function	51
3.2.6	Spirometry	52
3.3	'Omics' datasets	52
3.3.1	Metabolomics	52
3.3.2	Glycosylation of Immunoglobulin G	57
3.3.3	Microbial sequencing	63
3.3.4	Whole genome sequencing	64
3.4	Statistical analysis	65
3.4.1	Heritability	65
3.4.2	Univariate analysis	67
3.4.3	Pathway enrichment	69
3.4.4	Multi-variable regression (LASSO)	69
3.4.5	Imputation	71
3.4.6	Assessing predictive performance	71
3.4.7	Graphical models	73
3.4.8	Mixed Graphical Models	74
3.4.9	Graph analysis	75
4	Metabolic markers of leukocyte telomere length	77
4.1	Methods	77
4.1.1	Leukocyte telomere length	78
4.1.2	Metabolomics measurements	78
4.1.3	Ageing phenotypes	78
4.1.4	Gene expression	79
4.1.5	DNA methylation	79
4.1.6	Statistical analysis	79
4.2	Results	80
4.2.1	Independent associations with LTL	82
4.2.2	Ageing phenotypes	82
4.2.3	Expression of enzymes PLA2 and GGT	82
4.2.4	Smoking-related DNA de-methylation	84
4.2.5	Integration of results	84
4.3	Discussion	86
4.3.1	Lysolipids	86

4.3.2	Gamma-glutamylamino acids	87
4.3.3	4-Vinylphenol sulfate	87
4.4	Conclusion	88
5	Glycosylation profile of IgG in moderate kidney dysfunction	89
5.1	Methods	90
5.2	Results	91
5.3	Discussion	96
5.3.1	Galactosylation of IgG	96
5.3.2	Sialylation	96
5.3.3	Bisecting N-Acetylglucosamine	97
5.3.4	Genetics of kidney disease	97
5.3.5	Aiding diagnosis of CKD	98
5.4	Conclusion	98
6	Metabolic markers of renal disease in type 2 diabetics compared to non-diabetics	99
6.1	Methods	100
6.1.1	Study populations	100
6.1.2	Metabolomic profiling	101
6.1.3	Phenotypes	101
6.1.4	Statistical analysis	101
6.2	Results	103
6.2.1	Common markers of renal function	104
6.2.2	Markers of renal function in diabetics	106
6.2.3	Markers of renal function in non-diabetics	106
6.2.4	Metabolites differently associated with eGFR	106
6.3	Discussion	111
6.3.1	Renal production of amino acids	111
6.3.2	Energy-related metabolites	111
6.3.3	Lipoprotein composition	112
6.3.4	Phospholipids in HDL	113
6.4	Conclusion	113
7	The faecal metabolome as a functional readout of the gut microbiome	115
7.1	Methods	116
7.1.1	Study population	116
7.1.2	Data collection	116
7.1.3	Statistical analysis	118
7.2	Results	120
7.2.1	Faecal metabolite associations with phenotypes	120
7.2.2	Host genetic influence on the faecal metabolome	121
7.2.3	Microbial association with the faecal metabolome	124
7.3	Discussion	125
7.3.1	Faecal metabolites are stable during ageing	125
7.3.2	The faecal metabolome is influenced by host genetics	128
7.3.3	The faecal metabolome reflects microbial composition	128

7.3.4	Faecal metabolites associate with obesity	129
7.3.5	Faecal metabolites associated with renal function	129
7.4	Conclusion	131
8	Differential multi-fluid networks identify processes involved in end-stage renal disease	132
8.1	Methods	133
8.1.1	Study populations	133
8.1.2	Metabolomics measurements and normalisation	133
8.1.3	Statistical analysis	135
8.2	Results	136
8.2.1	Univariate differences between cases and controls	136
8.2.2	Multi-fluid graphical models	137
8.2.3	Network differences	137
8.3	Discussion	143
8.3.1	Salivary branched-chain amino acids	143
8.3.2	Steroid metabolism	144
8.3.3	Xanthine metabolism	145
8.4	Conclusion	146
9	Exploring the molecular basis of age-related disease comorbidities	148
9.1	Methods	149
9.1.1	Study Population	149
9.1.2	Data Acquisition and Processing	149
9.1.3	Network Inference	152
9.1.4	Network Analysis	152
9.1.5	Network Stability	152
9.2	Results	153
9.2.1	Topological properties of Age-MGM	153
9.2.2	Modularity of the Age-MGM	154
9.2.3	Robustness of the Age-MGM	155
9.3	Discussion	157
9.3.1	Decline of renal function links age with metabolic shift	157
9.3.2	Urate mediates association of renal function with body composition	157
9.3.3	Lung function as a central ageing process	158
9.3.4	Hormone expression directly associated with body composition	159
9.3.5	IgG glycosylation as new mechanism of obesity-associated inflammation	159
9.3.6	IgG-mediated inflammation associated with renal function	160
9.3.7	RBM20 as mediator of dyslipidaemia with advancing age	160
9.3.8	SEL1L2 associated with IgG glycosylation	160
9.3.9	Association of low meat diet and renal function mediated by CMPF	161
9.3.10	Adaption of cell membrane to alcohol intake	161
9.3.11	NCAM2 affects body composition	161
9.4	Conclusion	162

10 Conclusion and future directions	163
10.1 Summary of findings	163
10.1.1 Kidney disease	165
10.1.2 Biological ageing	166
10.2 Limitations	166
10.2.1 Systems biology	166
10.2.2 Causality	168
10.2.3 Translation to clinical application	168
10.3 Other confounding factors	169
10.4 Future directions and lessons learnt	170
Bibliography	171
Appendix	192
A Metabolic markers of telomere length	193
B Glycosylation profile of IgG in moderate kidney dysfunction	198
C Metabolic markers of renal disease in type 2 diabetics compared to non-diabetics	201
D The faecal metabolome as a functional readout of the gut microbiome	211
E Differential multi-fluid networks identify processes involved in end-stage renal disease	220
F Exploring the molecular basis of age-related disease comorbidities	235
G Published Articles	242
G.1 Integration of 'omics' data in aging research: from biomarkers to systems biology	243
G.2 Metabolomics profiling reveals novel markers for leukocyte telomere length	256
G.3 Glycosylation Profile of IgG in Moderate Kidney Dysfunction	267
G.4 Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model	277

List of Figures

1.1	The structure of telomeres	22
1.2	Schematic structure of a kidney	24
1.3	Genetic regulation of kidney function	26
1.4	Interdependencies of 'omics' layers	34
1.5	Graphs	36
1.6	Spurious correlations	40
3.1	Population characteristics of the TwinsUK cohort	53
3.2	Targeted and untargeted metabolomics	53
3.3	Metabolic pathways covered by Metabolon platforms	56
3.4	Exemplary NMR spectrum	58
3.5	NMR metabolomics provided by Brainshake Ltd.	59
3.6	The structure of Immunoglobulin G	60
3.7	Immunoglobulin G glycan chromatogram	60
3.8	Reproducibility of IgG glycan measurements	63
3.9	Twin modelling	67
3.10	Assessment of statistical predictions	72
3.11	Topological properties of nodes in graphs	76
4.1	Telomere length in TwinsUK and KORA	78
4.2	Prediction of leukocyte telomere length	83
4.3	Metabolic signature of telomere length	84
4.4	Phospholipase A	86
5.1	IgG glycosylation markers associated with renal function	93
5.2	Using IgG glycans to predict CKD	94
5.3	IgG glycosylation in twins discordant for CKD	95
6.1	Flowchart illustrating the identification of metabolic markers of diabetic and non-diabetic renal disease	102
6.2	Comparison of metabolite associations with renal function between type 2 diabetics and non-diabetics	105
6.3	Metabolites associated with eGFR in diabetic and non-diabetic cohorts	107
7.1	Study population	117
7.2	The frequency of faecal metabolites in TwinsUK	117

7.3	Faecal metabolites differ between the young and the elderly	121
7.4	Variance components of the faecal metabolome	122
7.5	Manhattan plot of faecal metabolites	123
7.6	Associations of faecal metabolites with gut microbes	126
7.7	Integrating faecal metabolites and microbes	127
7.8	Association of faecal amino acids with visceral fat	130
8.1	Metabolites detected in plasma, urine and saliva	134
8.2	Metabolites differing between CKD cases and controls	136
8.3	Network properties	138
8.4	Differential metabolic multi-fluid network	142
8.5	Caffeine Metabolism	145
9.1	‘Omics’ datasets in the TwinsUK cohort	149
9.2	Age-MGM variable selection and inference process	151
9.3	Ageing graphical model	155
9.4	Ageing network modules	156
A.1	Meta-analysis of metabolites associated with telomere length	197
C.1	Absolute concentrations of cholesterol	201
D.1	The effect of storage time on faecal metabolites	211
D.2	Variance components of the faecal metabolome	212
D.3	Faecal metabolic traits associated with host genetics	213
D.4	Genetic loci associated with faecal metabolic traits	214
D.5	Regional association plot of faecal metabolites	215
E.1	Comparison of renal function associations with previous studies	232
E.2	Principal components of metabolomics measurements	233
E.3	Metabolites associated with CKD across fluids	234
E.4	Rare metabolites differing between CKD cases and controls	234
F.1	Pairwise correlations in the Age-MGM	238
F.2	Connections within and between groups of variables of the Age-MGM	238
F.3	Stability of Age-MGM with respect to edge cut-off	239
F.4	Stability of the Age-MGM across datasets	240

List of Tables

1.1	Chronic kidney disease	25
1.2	Ageing studies	28
3.1	Population characteristics	46
3.2	Population characteristics of the TwinsUK cohort	47
3.3	IgG glycan traits	61
4.1	Population characteristics	80
4.2	Metabolites significantly associated with leukocyte telomere length	81
4.3	Associations of telomere length and metabolites with age-related phenotypes	85
5.1	Glycomics marker of renal function: Population characteristics	90
5.2	IgG glycans significantly associated with eGFR	92
6.1	Metabolic markers of renal function: population characteristics	102
6.2	Metabolites differently associated with eGFR in diabetics and non-diabetics	104
6.3	Metabolites associated with renal function	108
7.1	Population characteristics	117
7.2	Genetic loci associated with faecal metabolites	124
8.1	Population characteristics	135
8.2	Cross-fluid edges	138
8.3	Properties of the multi-fluid GGMs	138
8.4	Differential edges	139
A.1	Associations of circulating metabolites with telomere length in TwinsUK cohort	193
B.1	Associations of IgG glycan traits with renal function	198
B.2	Associations of whole plasma glycan traits with renal function	199
B.3	Comparison of renal-associated IgG glycosylation profiles between UPLC and MS/MS measurements	200
C.1	Associations of NMR metabolites with microvascular diseases in the GenoDi-abMar cohort	202
C.2	Longitudinal associations of metabolites with eGFR	204
C.3	Associations of NMR metabolite with renal function	206

D.1	Variance components and phenotype associations of the faecal metabolome	217
E.1	Metabolites included in the differential model	220
F.1	Omics variables included in the Age-MGM	235
F.2	Phenotypes included in the Age-MGM	236
F.3	Principal component-derived dietary patterns	237
F.4	Stability of the Age-MGM against variable selection procedure	241

Glossary

AD	Alzheimer's disease
ADCC	antibody-dependent cell-mediated cytotoxicity
aDMR	ageing-related differentially methylated region
AIC	Akaike information criterion
ANCA	anti-neutrophil cytoplasmic autoantibodies
AUC	area under the curve
BCAA	branched-chain amino acid
BIC	Bayesian information criterion
BMC	bone mineral content
BMD	bone mineral density
BMI	body mass index
BMIQ	beta mixture quantile dilation
BN	Bayesian Network
bp	base pair
CIG	conditional independence graph
CKD	chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CN	correlation network
COPD	chronic obstructive pulmonary disease
CPSS	complementary pairs stability selection
CT	computed tomography
CVD	cardiovascular disease
DAG	directed acyclic graph
DBP	diastolic blood pressure
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DXA	dual-energy X-ray absorptiometry
DZ	dizygotic
eGFR	estimated glomerular filtration rate
ESRD	end-stage renal disease
FDR	false discovery rate
FEV1	forced expiratory volume in one second
FFQ	food frequency questionnaire
FN	false negative

FP	false positive
FVC	forced vital capacity
FWER	family-wise error rate
GANI_MED	Greifswald Approach to Individualized Medicine
GC	gas chromatography
GFR	glomerular filtration rate
GGM	Gaussian graphical model
GM	graphical model
GRaFo	graphical random forest
GRN	gene-regulatory network
GSEA	gene set enrichment analysis
GWAS	genome-wide association study
HDN	human disease network
HILIC	hydrophilic interaction chromatography
IBD	inflammatory bowel disease
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KORA	Cooperative Health Research in the Region of Augsburg
LASSO	least absolute shrinkage and selection operator
LCL	lymphoblastoid cell line
LD	linkage disequilibrium
LOO	leave-one-out
LTL	leukocyte telomere length
MAF	minor allele frequency
MDRD	Modification of Diet in Renal Disease
MGM	mixed graphical model
MI	myocardial infarction
MIPS	Munich Information centre for Protein Sequences
MONICA	MONitoring trends and determinants in CARdiovascular disease
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSE	mean squared error
MuTHER	Multiple Tissue Human Expression Resource
MWAS	metabolome-wide association study
MZ	monozygotic
NHS	National Health Service
NMR	nuclear magnetic resonance
NPV	negative predictive value
OTU	operational taxonomical unit
PAGE	parametric analysis of gene set enrichment
PCA	principal component analysis
PCR	polymerase chain reaction
PLSDA	partial least squares discriminant analysis
PPI	protein-protein interaction
ppm	parts per million

PPV	positive predictive value
PRESS	predicted residual sum of squares
qPCR	quantitative polymerase chain reaction
REML	restricted maximum likelihood
RNA	ribonucleic acid
ROC	receiver operating characteristic
ROS	reactive oxygen species
rRNA	ribosomal RNA
RSD	relative standard deviation
RSS	residual sum of squares
SBP	systolic blood pressure
SD	standard deviation
SEM	structural equation modelling
SHIP	Study of Health in Pomerania
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SVM	support vector machine
T2D	type 2 diabetes
TN	true negative
TNR	true negative rate
TP	true positive
TPR	false positive rate
TPR	true positive rate
TSS	total sum of squares
UPLC	ultra-performance liquid chromatography
WGCNA	weighted gene co-expression network analysis
WGS	whole genome sequencing

Genes

ALT	<i>alanine aminotransferase</i>
ApoD	<i>apolipoprotein D</i>
ApoE	<i>apolipoprotein E</i>
ASPA	<i>aspartoacylase</i>
CYP	<i>cytochrome P450</i>
EPB42	<i>erythrocyte membrane protein band 4.2</i>
GGT	<i>gamma-glutamyltransferase</i>
GGT1	<i>gamma-glutamyltransferase 1</i>
GGTL3	<i>gamma-glutamyltransferase-like 3</i>
IgA	<i>immunoglobulin A</i>
IgG	<i>immunoglobulin G</i>

IL6	<i>interleukin-6</i>
ITGA2B	<i>integrin subunit alpha 2b</i>
LEPR	<i>leptin receptor</i>
NAT2	<i>N-acetyltransferase 2</i>
NCAM2	<i>neural cell adhesion molecule 2</i>
OXT	<i>oxytocin/neurophysin I prepropeptide</i>
PDE4	<i>phosphodiesterase 4</i>
PDE4C	<i>phosphodiesterase 4C</i>
PLA2	<i>phospholipase A2</i>
PLA2G15	<i>phospholipase A2 group XV</i>
PPAR- α	<i>proliferator-activated receptor α</i>
RARA	<i>retinoic acid receptor alpha</i>
RBM20	<i>RNA binding motif protein 20</i>
SEL1L	<i>SEL-1 suppressor of LIN-12-like 2</i>
SEL1L2	<i>SEL-1 suppressor of LIN-12-like</i>
SLC2A9	<i>solute carrier family 2, facilitated glucose transporter member 9</i>
ST6GAL1	<i>ST6 beta-galactoside alpha-2,6-sialyltransferase 1</i>
SVEP1	<i>sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1</i>
TRF1	<i>telomeric repeat binding factor 1</i>
TRF2	<i>telomeric repeat binding factor 2</i>

Metabolites

1,5-AG	1,5-Anhydroglucitol
2AB	2-aminobenzamide
ACN	acetonitrile
ADMA	asymmetric dimethylarginine
cAMP	cyclic adenosine monophosphate
CMPF	3-carboxy-4-methyl-5-propyl-2-furanpropionic acid
DHEA-S	dehydroepiandrosterone sulfate
EPA	eicosapentaenoic acid
FA	formic acid
GlcNAc	N-acetylglucosamine
GSH	glutathione
HDL	high density lipoprotein
IDL	intermediate density lipoprotein
LDL	low density lipoprotein
NAD ⁺	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)
PFPA	perfluoropentanoic acid
SCFA	short chain fatty acid
VLDL	very low density lipoprotein

CHAPTER 1

Introduction

Here I give an overview on ageing and ageing-related diseases, focusing on chronic kidney disease. I then review high-throughput 'omics' technologies and how they have contributed to ageing research. Finally, I introduce the concepts of systems biology and graphs and how these can be used to investigate mechanisms of age-related diseases.

Parts of this chapter have been published as review article in *Ageing Cell* (Zierer et al., 2015), which is attached in appendix G.1.

In contrast to Mendelian diseases, complex traits arise from the combined interplay of multiple genetic variants and environmental factors. All these variants usually have only small effect sizes and often affect several diseases simultaneously (Goh et al., 2007), which is referred to as pleiotropy. Due to pleiotropy and shared environmental risk factors comorbidities of diseases are a common phenomenon. Ageing is one of the most important risk factors for many diseases and thus promotes the occurrence of comorbidities.

The life expectancy in the UK increased by 5.3 years for men and 4.7 years for women over the last two decades and is predicted to further increase in the next twenty years (Office for National Statistics, 2014; Oeppen et al., 2002). With increasing life expectancy, age-related diseases are expected to also increase dramatically. One example is chronic kidney disease (CKD), which is the irreversible decline of renal function. In the USA, the prevalence of CKD has increased by 42 % between 1991 and 2004 (Coresh et al., 2007). Accordingly, the number of deaths attributed to CKD increased by 82 % between 1990 and 2010, CKD now being the 28th leading cause of death in the world (Lozano et al., 2012).

The increased prevalence of CKD has a major impact on public health care costs. For instance, the National Health Service (NHS) of England almost tripled its expenses on CKD patients from £566 million in 2002 to £1.4 billion in 2010, which by then constituted 1.3 % of its annual budget (Kerr et al., 2012). More importantly, health care for patients suffering from comorbidities of CKD, additional diseases caused by the impairment of renal function, was almost twice as expensive as for patients suffering from CKD only (Smith, 2004). The additional incidence of ≈ 7000 strokes and $\approx 12,000$ myocardial infarctions (MIs) due to CKD led to additional costs of about £175 million in England alone (Kerr et al., 2012). Therefore, finding ways to facilitate

healthy ageing and prevent disease onset and disease comorbidities is a major challenge of our time.

The development of high-throughput technologies facilitated the systematic investigation of the underlying biochemical pathways of complex traits. In contrast to traditional biochemistry, such techniques facilitate the unbiased analysis of disease pathways. Genome-wide association studies (GWASs) identified many genetic variants, mainly single nucleotide polymorphisms (SNPs), influencing age-related traits such as CKD (Köttgen et al., 2010; Okada et al., 2012) and telomere length (Codd et al., 2013). Besides genomics it is now possible to measure thousands of markers from various biological layers. Hundreds of ageing associations were found by analysing these new ‘omics’ datasets, such as epigenomics (Bell et al., 2012), transcriptomics (Glass et al., 2013), metabolomics (Menni et al., 2013b; Yu et al., 2012) and glycomics (Kristic et al., 2014), improving our understanding of ageing and its effect on disease susceptibility.

However, the different ‘omics’ layers are highly correlated within and between each other. For example, metabolite levels correlate with each other because they are linked by chemical reactions (Krumstiek et al., 2011), which are usually catalysed by proteins. Common genetic variation that affects the efficacy or abundance of these proteins has been shown to strongly influence circulating metabolite levels (Shin et al., 2014). The resulting increase or decrease in metabolite levels can cause one or possibly multiple diseases, which then occur together due to their shared underlying cause.

Considering all these dependencies is crucial to obtaining a mechanistic understanding of complex phenotypes. While association studies are very successful in identifying the relevant components of a process, systems biology aims to analyse their interactions and multivariate interdependencies (Barabási et al., 2004; Ideker et al., 2001). The increased availability of large ‘omics’ datasets has facilitated the propagation of integrative systems biology approaches.

The rest of this chapter is divided into three sections: First, I will outline the hallmarks of ageing and CKD, which will serve as an example for age-related diseases in this thesis. Then, I will review the most commonly used ‘omics’ technologies and discuss their impact on ageing research. Finally, I will describe the principles of systems biology, again focusing on their application in the field of ageing.

1.1 Ageing

Ageing is often described as the progressive accumulation of changes with time. In biological systems, these changes lead to a loss of physiological aptitude and fertility, an increased susceptibility to disease, and ultimately to death (Harman, 1988; Kirkwood et al., 2000; Vijg et al., 2005; López-Otín et al., 2013; Harman, 2001). Despite considerable effort and the development of many theories, the underlying processes of ageing are still largely unknown (Rattan, 2006; Kirkwood et al., 2000; Weinert et al., 2003).

1.1.1 Theories of ageing

More than 300 theories exist that attempt to explain why and how ageing occurs (Viña et al., 2007). Most of these focus on a single cause of ageing. However, ageing is a multi-factorial process and likely a combination of theories is true (Weinert et al., 2003).

Many theories consider ageing a consequence of evolution. Ageing – and ultimately death – can thereby be considered a programmed feature of evolution to maintain the fitness of the entire population by removing the elderly. On the other hand, ageing can be a consequence of the lack of evolutionary pressure after replication (Longo et al., 2005). Also, evolution might favour mutations that increase health in younger age, even if they are deleterious in later age (Kirkwood et al., 2000).

Multiple mechanisms were proposed on how ageing takes place. The wear and tear theory of ageing suggests that ageing is a consequence of the deterioration of cells due to their metabolic activity. The free radical theory of ageing proposes that the accumulation of reactive oxygen species (ROS) is the cause of increased stress, slowly causing damages the DNA and other parts of the cell (Harman, 1956). Several extensions of this theory consider the mitochondria as main source of ROS, due to malfunction of the respiratory chain, as well as the main target of ROS, as mitochondrial DNA is not as well protected and easily repaired as nuclear DNA (Bratic et al., 2013). The theory of ‘inflammageing’ suggests that increasing chronic low-grade inflammation, as commonly observed in elderly, might be a shared root cause of age-related diseases, hence causing their joint occurrence (Franceschi et al., 2014).

Other potential causes of ageing are the accumulation of somatic mutations due to errors in DNA replication (Freitas et al., 2011) and the decreasing accuracy of gene expression (Weinert et al., 2003). These changes could lead to decreased RNA, DNA, and cell turnover, which in turn decreases the ability of the body to react to environmental stress and repair injuries. A special case of a DNA damage theory of ageing is the telomere theory of ageing. Telomeres form the end of each chromatid and, due to the inability of the DNA polymerase to replicate the 3' end of the template strand telomeres, shorten with every cell cycle. Thus, telomeres have been proposed as a cause of ageing (see section 1.1.2.1 for details). The shortening of telomeres is one reason for chromosome instability, which is another common feature of ageing (Blasco, 2007).

A myriad of studies provides evidence for the shortening of telomeres with age (Harley et al., 1990), the involvement of ROS in ageing (Beckman et al., 1998), changes of gene expression (section 1.2.3), and many other theories of ageing. However, differentiating between causes and consequences of ageing remains a major challenge. Many of the proposed mechanisms might even be cause and consequence simultaneously, thus mutually reinforcing each other in a vicious circle. For instance, increased abundance of ROS causes (mitochondrial) DNA damage, which causes weakening of anti-oxidant mechanisms and damage to the respiratory chain, which causes further increase of ROS (De Grey, 2005).

1.1.2 Biological ageing

Due to the increased interest in healthy ageing, many studies investigated the physiological changes that occur with age and how these make us more susceptible to disease. However, chronological age, which is the time since birth, was shown to be a bad predictor for health and fitness (McClintock et al., 2016). Thus, researchers came up with the concept of biological age, which takes into account an individual's actual physical and mental health, aiming to capture personal differences in the process of ageing. While chronological age is easy to measure, there is no consistent measure of biological age. Instead, different proxies are used to quantify differences in biological age.

One way to measure biological ageing is to assess an individual's physiological capacities. Various tests measuring – amongst others – grip strength, lung function and power are available for epidemiological studies. To assess physical and psychological aptitude more comprehensively, frailty indices (section 1.1.2.2) combine multiple such measures with information on several diseases as well as biochemical measures. Another measure of biological ageing is telomere length, which is thought to be one cause of ageing (see section 1.1.1). By studying centenarians, researchers investigate successful, i.e. slow, ageing while avoiding to explicitly define biological age.

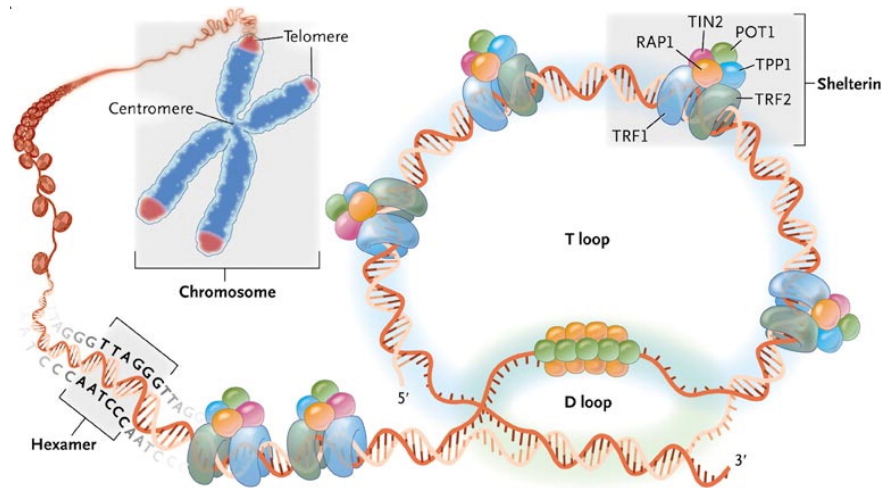
All these proxies of biological ageing were used in various studies to research potential causes and consequences of ageing. However, the ultimate goal of most ageing research is to prevent or cure disease to facilitate healthy ageing. Consequently, many studies focus on individual diseases rather than these more abstract measures of ageing. In the following, I will outline some of the above-mentioned ways to quantify biological ageing in greater detail. I will, then, give an overview on CKD, which is used in this thesis exemplarily for age-related diseases (section 1.1.3).

1.1.2.1 Telomere length

Leukocyte telomere length (LTL) is a measure of biological age, that is based on the theory of telomere shortening (Mather et al., 2011; Harley et al., 1992). Telomeres are repetitive DNA sequences that are located at the end of each chromatid. Together with proteins that specifically bind to telomeres, such as *telomeric repeat binding factor 1* (TRF1) and *telomeric repeat binding factor 2* (TRF2), they form large nucleoprotein complexes at the ends of the chromosomes. These so-called t-loops (De Lange, 2005) protect the DNA from degradation and end-to-end fusion (Figure 1.1).

During DNA-replication, the DNA polymerase binds to RNA primers that are synthesised on the template strand by the primase and elongates the new DNA strand from the 5' to the 3' end, which corresponds to the 3' to 5' direction of the template strand. As the primer has nowhere to bind upstream of the 3' end of the template, the very end of the chromosome cannot be replicated. Thus, the newly synthesised DNA strand will be missing some nucleotides at its very 5' end and thus, telomeres shorten with every cell cycle (Levy et al., 1992; Allsopp et al., 1995a). The continuous shortening of telomeres will continue throughout ageing until the telomeres become too short to further protect chromosomes. Cells that reach this so-called Hayflick limit (Olovnikov, 1996) stop dividing and switch to a phase of senescence (Allsopp et al., 1995b).

Figure 1.1 The structure of telomeres. Telomeres are the repetitive DNA sequences that make up the last ≈ 10 kbp of each chromatid. Together with several DNA-binding proteins they form the t-loop to protect the ends of the chromosomes from deterioration and end-to-end fusion. (Reproduced with permission from Calado et al. (2009), Copyright Massachusetts Medical Society)



While this process affects all cells equally, the enzyme telomerase elongates telomeres in germ lines to prevent their ageing. However, this enzyme is inactive in somatic cells (Collins et al., 2002). The shortening of telomeres has been proposed as ‘mitotic clock’, limiting the replicative life span of cells and causing cell senescence (Harley et al., 1992).

LTL has been used as proxy for biological age in a multitude of ageing studies and has been found to be strongly associated with chronological age (Valdes et al., 2005). Additionally, it was found to be significantly associated with several age-related diseases, including CKD (Harst et al., 2008; Raschenberger et al., 2015), Alzheimer’s disease (AD) (Panossian, 2003; Thomas et al., 2008), cardiovascular disease (Fitzpatrick et al., 2007; Brouillette et al., 2003), osteoarthritis (Zhai et al., 2006; Harbo et al., 2012) and cancer (Shay et al., 2011; Artandi et al., 2009), independently of chronological age. LTL was also shown to predict mortality (Cawthon et al., 2003; Kimura et al., 2008) and longevity (Vera et al., 2012).

Telomere length is, thus, a powerful biomarker of ageing, which combines the theoretical understanding of its relationship with ageing with the empirical evidence that relates LTL with diseases. Many studies aimed to further explore the mechanisms that cause telomere shortening and subsequent health deficits. For instance, GWASs have successfully identified a number of genes associated with LTL (Codd et al., 2010; Codd et al., 2013), including TRF1 and TRF2. However, combined they only account for 1.2 % of the observed variance and it is still poorly understood how the shortening of telomeres affects an individual’s health. In chapter 4 I will further explore such potential mechanisms using metabolic profiling.

1.1.2.2 Frailty

The concept of frailty aims to describe the inability of a person to live an independent and active life. It includes power and muscle mass as well as the dependency on others, particularly hospital care (Rockwood et al., 1994). While there is no strict definition of frailty, a common concept to quantify frailty are frailty indices, such as the Rockwood frailty index (Mitnitski et al., 2001; Rockwood, 2005). These indices count the proportion of health deficits an individual is suffering from. To this end, usually many clinical and sub-clinical tests are combined, including symptoms, such as sleeping problems, clinical chemistry measures, such as elevated creatinine levels, and diseases, such as type 2 diabetes (T2D) (Mitnitski et al., 2001). Therefore, frailty indices are very general markers of health and studies on frailty aim to investigate the underlying causes that lead to healthy ageing rather than identify biomarkers of specific diseases.

1.1.2.3 Longevity

Longevity, living an exceptionally long life, has been a topic of interest for most of human history. It was found to be about 20 % heritable, though estimates vary considerably between studies (Murabito et al., 2012). Centenarians tend to either escape age-related diseases, such as dementia and cancer, or survive them, such as cardiovascular diseases (Arnold et al., 2010). Consequently, centenarians are examples of successful, healthy ageing and have been studied for several decades to explore causes of the delayed biological ageing (Willcox et al., 2010).

The first genetic study on longevity compared centenarians with younger controls and identified the only gene that has since been consistently linked with longevity: the *apolipoprotein E* (ApoE) gene (Schächter et al., 1994; Deelen et al., 2011). A later study achieved to distinguish centenarians from younger controls with an 89 % sensitivity and specificity using a multivariate genetic model incorporating 281 SNPs, indicating that longevity is a complex trait (Sebastiani et al., 2012). Studying the metabolism of centenarians revealed that they are in some aspects more like young individuals than to the elderly (Collino et al., 2013). In particular, markers of lipid peroxidation, i.e. oxidative stress, were low in centenarians and young individuals, but high in the elderly.

While centenarians are very useful to study longevity, there is a natural limit in sample size, which for instance prevents larger GWASs. Also, lifestyle factors, which differ between age groups, might affect the results of centenarian studies.

1.1.3 Chronic kidney disease

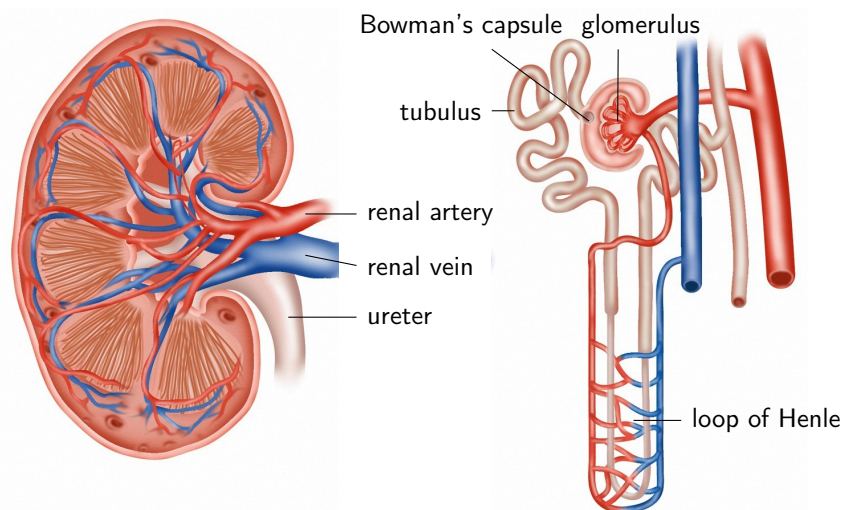
Chronic kidney disease (CKD) is a common disease in the Western world (Table 1.1). As ageing is the most important risk factor for renal disease, the prevalence of CKD is increasing due to the ageing populations of developed countries (Tonelli et al., 2014; Prakash et al., 2009). Besides age, obesity, diabetes and hypertension are major risk factors for developing CKD (Hsu et al., 2006; Lea et al., 2002). Thus, the global obesity epidemic causes further increase of the prevalence of CKD. Patients suffering from CKD do not only become dependent on dialysis

or renal transplant once they progress to end-stage renal disease (ESRD), but are also much more likely to experience cardiovascular events, such as MI (Sarnak et al., 2003). This led to the loss of 928,000 lives and 14,754,000 disability-adjusted life years worldwide in 2004 (Ayodele et al., 2010). There is no cure for CKD, hence early diagnosis and prevention are essential. In this section I will first give an overview over the physiology of the kidneys and then outline the hallmarks of CKD.

The kidneys are two organs located in the back of the abdominal cavity left and right of the spine, which are mainly responsible for the filtration of blood. About 20 % of the blood leaving the heart is pumped through the kidneys where it is filtered and waste products are excreted as urine (Boron et al., 2012). To this end each kidney consists of about 1,000,000 nephrons, which in turn consist of the glomerulus and the tubule (Figure 1.2). Due to the high blood pressure in the glomerulus, which is the first part of each nephron, water and solved compounds are filtered through a semi-permeable membrane and absorbed by the Bowman's capsule. From the capsule the filtrate is transported through the tubule, where some solutes are reabsorbed in the blood and others are further excreted. Finally, the filtrate is transported to the urinary bladder while the filtered blood leaves the kidneys through the renal vein. The amount of fluid that is filtered by the glomeruli per time is referred to as glomerular filtration rate (GFR) (Boron et al., 2012).

As an individual ages, the kidneys become less efficient in filtrating blood, which results in the accumulation of toxins in the blood as well as unintended excretion of proteins in the urine – proteinuria. This decrease of renal filtration is referred to as chronic kidney disease (CKD). CKD is a diverse disease that arises due to a complex interplay of genetics and environmental

Figure 1.2 Schematic structure of a kidney. The kidney is a bean-shaped organ that is supplied with blood via the renal artery. Each kidney consists of about 1 million nephrons (depicted on the right side), which filter the blood in the round glomerulus and along the tubulus. Filtered blood leaves the kidneys through the renal vein, while the filtrate is excreted through the ureter to the bladder. (adapted with permission from Wellcome Images, <https://wellcomeimages.org>)



causes. It can manifest a range of symptoms, ranging from mild, non-symptomatic renal damage to complete renal failure and eventually death (Levey et al., 2012). It is thus divided into several stages based on the GFR, where $\text{GFR} > 90 \text{ mL/min/1.73m}^2$ corresponds to stage 5 (renal healthy) and $\text{GFR} < 15 \text{ mL/min/1.73m}^2$ to stage 1 (end-stage renal disease, ESRD) (c.f. Table 1.1). Other definitions also consider the excretion of protein – usually albumin – in the urine to define stages of CKD (Coresh et al., 2007).

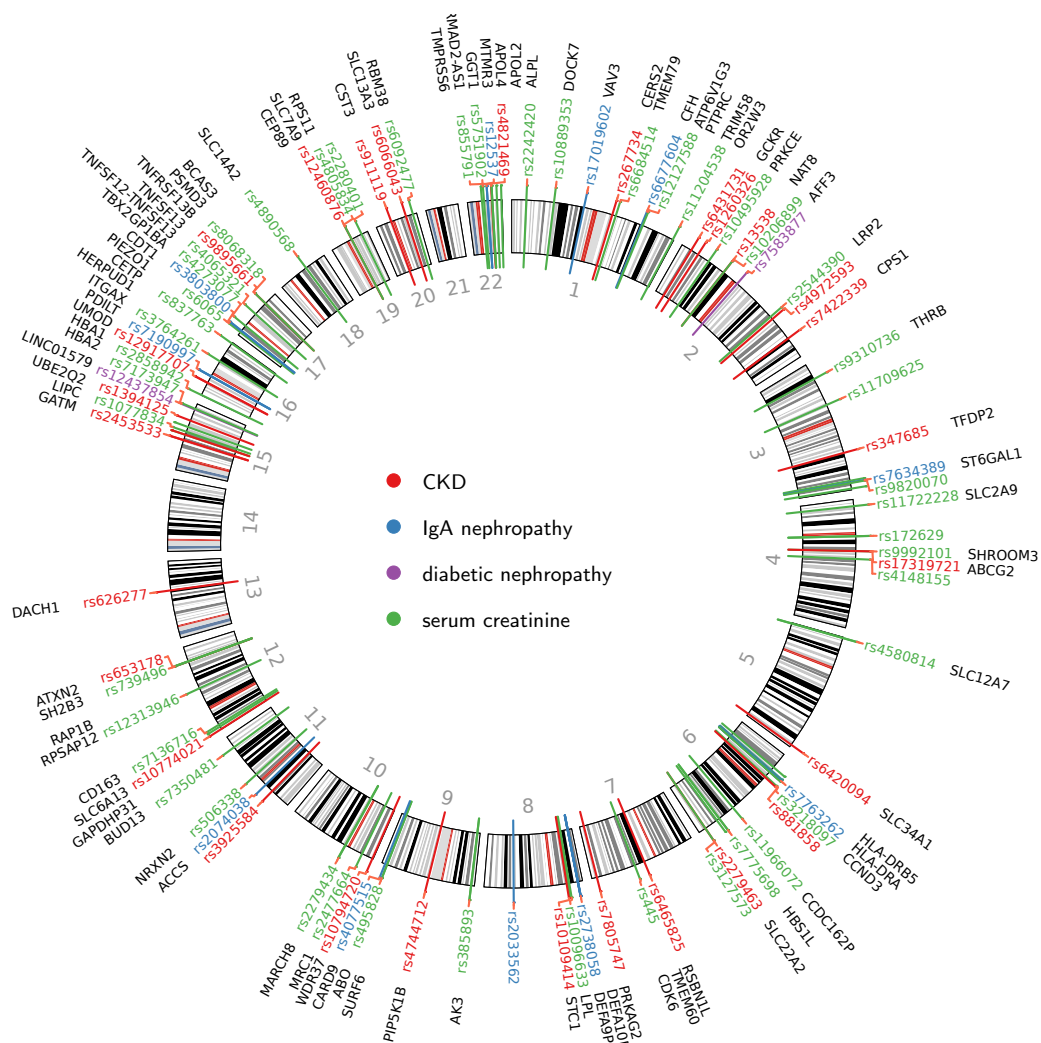
Timely diagnosis of renal damage remains difficult. Precise measurement of the GFR requires intravenous administration of inulin, which is rarely done in clinical practice or research (Perrone et al., 1992). Instead, the GFR is most commonly estimated based on the concentration of creatinine in serum (Perrone et al., 1992; Levey et al., 2012). Creatinine is produced by the muscles, secreted into the blood and then filtered by the kidneys where it is excreted unchanged in urine. Creatinine provides the advantage of straight-forward, fast and minimally-invasive measurements by routine blood test. However, creatinine is an insensitive marker and its concentration might start to increase only when about half of the renal capacity is already lost (Stevens et al., 2009), thus hampering early diagnosis (Thomas et al., 2009). Also, creatinine concentrations vary according to age, gender, race. All these factors must be considered to calculate the estimated glomerular filtration rate (eGFR) from creatinine (see section 3.2.3 for details).

To improve the diagnosis and treatment of CKD, many studies investigated physiological changes that are associated with renal function and showed that CKD is heritable with heritability estimates ranging between 0.33 and 0.41 (Langeveld et al., 2004; Bochud et al., 2005). Large-scale GWASs investigated the genetic architecture of renal function in cohorts of up to 175,000 individuals and found more than 100 genetic loci associated with renal function (Köttgen et al., 2010; Pattaro et al., 2012; Pattaro et al., 2016) (Figure 1.3). However, effect sizes of these loci are small and even all of them combined explain only 3.2 % of the observed phenotypic variance (Pattaro et al., 2016). Besides genetics, renal function is strongly affected by environmental factors, including obesity and smoking (Prakash et al., 2009). Decline of renal function is also accompanied by drastic changes in the metabolism, some of which could potentially be used as alternative markers for renal function in the future (Sekula et al.,

Table 1.1 Chronic kidney disease. CKD is commonly divided into five grades according to the capabilities of the kidneys to filter the blood, which is quantified by the glomerular filtration rate (GFR). Prevalence of different stages of CKD are based on the NHANES survey published by Coresh et al. (2007).

GFR (mL/min/1.73m ²)	CKD stage	Prevalence
>90	healthy	40.7 %
89 to 60	2 mildly reduced renal function	51.2 %
59 to 30	3 moderately reduced renal function	7.7 %
15 to 29	4 severely reduced renal function	0.4 %
<15	5 end-stage renal disease	

Figure 1.3 Genetic regulation of kidney function. Chronic kidney disease (CKD) and more generally renal function is strongly influenced by genetics. Here I summarise the results of 21 genome-wide association studies (GWASs) on different aspects of renal function, listed in the GWAS Catalogue (Welter et al., 2014, version 1.0.1). They identified a total of 131 SNPs in 119 unique genes. Genome-wide significant associations ($p < 5 \times 10^{-8}$) were retrieved for four renal traits: CKD (keywords: '*chronic kidney disease*'), IgA nephropathy ('*IgA glomerulonephritis*' or '*kidney disease*'), diabetic nephropathy ('*diabetic*' and '*nephropathy*') and concentration of creatinine in serum ('*serum creatinine*'). SNPs are coloured according to the associated trait.



2016). However, the causal direction of most of these associations remains elusive, and many of them might be consequences of the incomplete filtration of blood rather than causes of CKD.

1.2 ‘Omics’

Since the first sequencing of the human genome in 2001 (Lander et al., 2001), an increasing number of high-throughput technologies has become available. These aim to provide holistic measurements of the entire variation of the genome, transcriptome, metabolome and others. Such ‘omics’ technologies provide valuable tools to studying ageing and related diseases on the molecular level. Numerous association studies identified epigenetic mutations, gene expression levels, metabolite concentrations, etc. correlated with chronological and biological age (Table 1.2) (Valdes et al., 2013). In the following, I will describe the most commonly used ‘omics’ technologies – genomics, epigenomics, transcriptomics, proteomics, metabolomics and microbiomics – focusing on their application in ageing research. Methodological details of the technologies used in this thesis will be outlined in chapter 3.

1.2.1 Genomics

Genomics was the first omics field for which high-throughput measurements became available due to the development chips for genotyping. Current chips can measure up to 5 million SNPs (Ha et al., 2014). Today, next-generation sequencing technology is slowly replacing the chip technology as the cost of sequencing has dropped below \$0.1 per million base pairs (bps) (Liu et al., 2012). Thus, gene variation is nowadays often available at single nucleotide resolution.

While ageing (or rather longevity) itself was found to be only about 20 % heritable (Murabito et al., 2012), many age-related diseases such as AD have high heritability >70 % (Gatz et al., 2006) and others like osteoarthritis (Ishimori et al., 2010) or cataract have a heritability of 50 % (Hammond et al., 2001). The GenAge database contains about 300 genes which are thought to be related to ageing based on homology with model organisms such as nematodes or mice (Tacutu et al., 2013). Sebastiani et al. (2012) developed a model containing 281 SNPs to distinguish between centenarians and younger controls in a cohort of 1715 people. However, the only age-related gene that has been reliably replicated is ApoE. Common genetic variants at this locus have been associated with accelerated ageing and cognitive decline (Johnson, 2006; Davies et al., 2014), possibly because it increases the risk for coronary artery disease, stroke and AD (Smith, 2002).

1.2.2 Epigenomics

Epigenomics describes the study of heritable changes of the genome that are not caused by mutations of the DNA sequence. The most commonly studied epigenetic mechanism is DNA methylation, which is known to often but not always silence gene expression. In contrast

Table 1.2 Ageing studies. Numerous 'omics' studies have been conducted to investigate molecular changes associated with ageing (see section 1.2). This table provides an overview over studies across 'omics' fields.

	Platform	Variables	Samples	Main finding	Ref.
Epigenetics	Illumina 27K	26,690	172	490, predominantly hyper-methylated, DMRs associated with chronological age and one DMR, in the <i>TBX20</i> gene also associated with age-related phenotypes	Bell et al. (2012)
	Illumina 27K and 450K	21,369	8000 from 82 datasets	age-related methylation changes are stable across various tissues and even species; tissue ageing can be measured using 353 CpG sites	Horvath (2013)
Transcriptomic	Illumina 27K	102	256 from four datasets	only three CpG sites are sufficient to reliably predict age	Weidner et al. (2014)
	Affymetrix HG-U95Av2	12,000	30	463 differentially expressed genes in brain tissue	Lu et al. (2004)
	Affymetrix Hg-U133plus 2.0	47,000	58	thousand of genes in four brain regions differentially expressed with age; these overlap between different brain regions	Berchtold et al. (2008)
	Illumina Human HT-12 V3 Bead-Chips	48,804	856	1672 genes in skin, 188 in adipose tissue, and 2 in LCL correlate with age; genes largely different between tissues	Glass et al. (2013)
Proteomics	two-dimensional difference gel el	2700	18	higher abundance of proteins involved in aerobic metabolism and a lower abundance of proteins involved in anaerobic metabolism in elderly	Gelfi et al. (2006)
	SOMALogic	1129	879	11 circulating proteins associated with chronological age and age-related phenotypes	(Menni et al., 2015)
Glycomics	UPLC	77	5117 from four cohorts	glycosylation of IgG strongly associated with chronological age	Kristic et al. (2014)
Metabolomics	untargeted GC-MS	>300	269	Widespread metabolic changes with age; e.g. lipid metabolites, such as long chain fatty acids and carnitines, lower concentrated in younger individuals	Lawton et al. (2008)
	targeted MS platform (Biocrates)	131	2886	strong and reproducible metabolic changes indicating incomplete mitochondrial fatty acid oxidation	Yu et al. (2012)
	untargeted MS platform (Metabolon)	260	6055	22 metabolites independently associated with age and age-related phenotypes explain 69 % of the variance	Menni et al. (2013a)
Microbiomics	16S sequencing	>40,000	170	microbiome relatively stable throughout life but changes drastically in late life	Claesson et al. (2011)
	Human Intestinal Tract Chip	129	84	significant changes in microbiome composition of centenarians, particularly decrease of alpha diversity	Biagi et al. (2010)

to the genome, which is the same in all cells, the epigenome is an important factor of cell differentiation leading to profound epigenetic differences across different cell types (Meissner, 2010).

The most often used methylation chip, the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA), measures over 450,000 methylation sites and covers 99 % of all RefSeq genes (Dedeurwaerder et al., 2011). However, it covers less than 10 % of variable regions (Ziller et al., 2013). The more recent Infinium MethylationEPIC BeadChip captures up to 850,000 methylation sites, but has not yet been used in ageing studies. Sequencing based methods, particularly bisulfite sequencing, facilitate the analysis of DNA methylation on a genome-wide level, but few large datasets are available due to its high cost (Ziller et al., 2014).

The epigenome is influenced by environmental and lifestyle factors (Alegría-Torres et al., 2011; Breitling et al., 2011) and is associated with many complex diseases such as neurodegenerative disorders (Portela et al., 2010) and cancer (Ehrlich, 2002; Horvath, 2013). In contrast to genomics, nearly 500 differentially methylated regions were found to be associated with chronological age and age-related phenotypes such as lung function, cholesterol levels and maternal longevity (Bell et al., 2012). A recent study by Weidner et al. (2014) showed that methylation patterns of just three sites are sufficient to reliably predict chronological age, thus suggesting that many of the previously identified methylation sites are not independently associated with age. Variations of methylation patterns are consistent across several tissues and cell types (Horvath, 2013) and together form a global pattern of hypomethylation in repetitive sequences, hypermethylation in promoter regions, and higher inter-cell variability (Cevenini et al., 2008; Bacalini et al., 2014).

There is also growing evidence for an epigenetic contribution to CKD (Reddy et al., 2015), which might even explain parts of its heritable component (Reddy et al., 2011). However, correlations between DNA methylation and eGFR are moderate (Wing et al., 2014; Smyth et al., 2014). Nonetheless, epigenomics is a promising tool in ageing research and its impact is expected to increase further with the availability of the larger EPIC chip and sequencing datasets.

1.2.3 Transcriptomics

Genes are transcribed into RNA molecules. The entirety of RNA transcripts is referred to as transcriptome. It can be divided in coding RNAs, which are further translated into proteins, and non-coding RNAs, which perform various functions, such as the regulation of gene expression (Eddy, 2001).

Similar to genetic and epigenetic variation, transcript abundances can be measured either by chips or sequencing methods. Gene expression was shown to dramatically change with age. A pioneer study comparing post-mortem human frontal cortex tissue samples of 30 individuals of different ages yielded 463 differentially expressed genes (Lu et al., 2004). Despite the small sample size, results could be replicated in subsequent experiments. Four years later, Berchtold et al. (2008) identified several thousand age-related changes of gene expression in four different

brain tissues. Later studies by different groups identified profound changes of the transcriptome with age in further tissues, such as skin, adipose tissue (N=865) (Glass et al., 2013) and the kidneys (N=134) (Rodwell et al., 2004). Interestingly, most of these changes did not overlap in different tissues. A meta-analysis across different species and tissues revealed only 73 genes consistently associated with age (Magalhães et al., 2009), suggesting that most observed age-related changes in the transcriptome are either species and tissue specific or false positive discoveries (Valdes et al., 2013). In the meta-analysis genes related to immune response and lysosome tended to be over-expressed, while genes related to mitochondria and oxidative phosphorylation were under-expressed in the elderly (Magalhães et al., 2009).

1.2.4 Proteomics

Proteins are translated from coding RNA transcripts. Due to alternative splicing and post-translational protein modifications, the number of proteins is estimated to be two orders of magnitudes higher than the number of genes. However, current proteomic techniques based on immunoassays, protein arrays or mass spectrometry (MS) can measure only a small fraction of the proteome (up to 1000 proteins in a sample). The most comprehensive description of the human proteome across various tissues to date consists of 18,097 proteins (19,376 isoforms) collected from ten thousands of MS experiments (Wilhelm et al., 2014). Due to these technicalities, 'proteomics' studies in ageing research has so far been focusing on smaller sets of proteins and small sample sizes.

In an early study of protein abundance in the vastus lateralis muscle, Gelfi et al. (2006) observed higher abundance of several proteins involved in aerobic metabolism and a lower abundance of proteins involved in anaerobic metabolism in the elderly. Besides this, six transport proteins were consistently under-expressed in older individuals. However, only 12 samples were analysed in this study without replication. A recent study by our group analysed over 1000 proteins in 200 plasma samples using the SOMAscan assay (Menni et al., 2015). 11 proteins were found to strongly associate with chronological age as well as age-related phenotypes, such as lung function and blood pressure, and the results were replicated in an independent cohort. Even though comprehensive proteomics studies are still missing, proteins are likely to be associated with several age-related diseases, as for instance cardiovascular disease (Mehra et al., 2005) and AD (Swardfager et al., 2010) are consistently associated with elevated levels of pro-inflammatory cytokines.

Several studies investigated the involvement of various proteins in CKD (Arthur et al., 2010; Mischak et al., 2015). However, due to the lack of high-throughput technologies in proteomics most of these investigated very limited number of proteins and/or based their assumptions on small sample sizes. Nevertheless, proteins are promising biomarkers of renal function. For instance, the presence of *albumin* in urine – albuminuria – is a common clinical marker for renal function. However, like creatinine it lacks sensitivity and specificity (Mischak et al., 2015). Another commonly used marker of CKD is *cystatin C* (Hojs et al., 2006), which predicts incident ESRD and all-cause mortality (Menon et al., 2007). There are several other, less well described proteins that have been associated with renal function, such as the *phospholipase A2* (PLA2) receptor (Beck et al., 2009). While these studies and many others provide insights in the pathology of CKD and suggest new tools for diagnosis, a holistic proteomics analysis of renal function is yet to be conducted.

1.2.5 Post-translational modifications - glycomics

Post-translational modifications are important elements of proteins, which can alter their biochemical properties such as protein structure, binding preferences, and enzyme activity. There are many different modifications ranging from the addition of small molecules (e.g. acetylation or phosphorylation), to the addition of larger molecules like lipid or sugar chains (e.g. palmitoylation, glycosylation), and addition of whole proteins (e.g. ubiquitination). The most common modification is glycosylation. The attached oligosaccharides – glycans – are thought to mainly serve as structural elements of proteins or specific binding sites for other glycans or proteins (Varki et al., 2009). However, glycans are highly diverse and many of them are not yet characterized or annotated. Thus, glycans might have many additional functions. For example, glycans in the gut act as food for microbes (Koropatkin et al., 2012), which could have immune functions that are important in ageing.

Recent development enable the high-throughput measurement of glycans of either a single protein or all proteins simultaneously (Pucic et al., 2011; Royle et al., 2008) (see section 3.3.2 for details). The application of this technology on epidemiological cohorts revealed that glycan structures are stable for one individual over time (Gornik et al., 2009) but very diverse within a population (Pucic et al., 2011; Knezević et al., 2009). Differences in glycomes were found to be related with various cancers (Adamczyk et al., 2012; Fuster et al., 2005).

Immunoglobulin G (IgG) is a particularly well-suited model for glycomics as its has well defined glycosylation sites and many studies investigated the effects of IgG glycosylation in great detail (Gornik et al., 2012). N-glycans attached to the conserved Asn297 in the Fc part of IgG are important modulators of the function of IgG (Jang et al., 2015), changing for instance its inflammatory potential. Recently Kristic et al. (2014) showed that IgG glycans are strongly associated with age: A linear combination of three glycans explained 58 % of the observed variance of chronological age in a study of four independent populations with 5117 participants in total. Using the same platform, I found glycosylation of IgG associated with renal function. The results of this study are presented in chapter 5.

1.2.6 Metabolomics

Metabolomics investigates the low-molecular-weight molecules in a biological system. The measured molecules are often referred to as metabolites since many of them act as educts, products, and intermediates of the cellular metabolism. Currently, the Human Metabolome Database (Wishart et al., 2013) contains more than 40,000 distinct metabolites from different tissues.

Like proteomics, to date, there is no analytical method available to determine and quantify all these metabolites in a single experiment. Current platforms, using either chromatography coupled with MS or nuclear magnetic resonance (NMR) spectroscopy, can measure up to roughly a thousand unique metabolites in untargeted settings and a smaller number using pre-defined targeted approaches. The restriction of the targeted approach comes with the advantages of higher sensitivity, absolute instead of relative quantification, and straight-forward compound identification (Menni et al., 2017b) (see section 3.3.1 for details).

In 2008, the first metabolome-wide association study (MWAS) on age analysed the plasma metabolome of 269 individuals using an untargeted approach. The authors found 100 out of 300 compounds significantly correlated with chronological age (Lawton et al., 2008). More recently, larger cohorts were employed to study the association of circulating metabolites and age using both targeted and untargeted metabolomics platforms. Yu et al. (2012) analysed 131 targeted metabolites in 2162 individuals from the Cooperative Health Research in the Region of Augsburg (KORA) cohort, while Menni et al. (2013b) analysed 280 untargeted metabolites in 6055 twins from the TwinsUK cohort. Both studies identified about half of the analysed metabolites to be associated with chronological age. Many of those metabolites were also found to significantly correlate with age-related phenotypes such as lung function, bone mineral density (BMD) and cholesterol levels (Menni et al., 2013b), AD (N=93) (Orešič et al., 2011), cancer (Teicher et al., 2012) and T2D (N=100) (Suhre et al., 2010; Menni et al., 2013a).

MWASs were also conducted to identify metabolic changes that accompany CKD and found, as expected, a large proportion of metabolites correlated with eGFR (Sekula et al., 2016; Mäkinen et al., 2012). In this thesis, I investigate these metabolic alterations in more depth, focusing on differences between diabetic and non-diabetic nephropathy (chapter 6) and cross-fluid changes (chapter 8).

1.2.7 Microbiomics

The human microbiome describes the complete set of microbial species (and their genomes) hosted by the human body. The largest microbial community resides in the gut, where microbial cells and their genes outnumber human cells (10:1) and genes (100:1) (Peterson et al., 2009; Zhu et al., 2010; Sender et al., 2016). More than 10,000 different species with millions of protein coding genes were identified by the Human Microbiome Project (Turnbaugh et al., 2007; Peterson et al., 2009; Biagi et al., 2012) and more than 1000 of these microbes have so far been fully sequenced.

Currently, microbiome studies use predominantly sequencing of the variable regions of the 16S gene to quantify the microbial community. 16S sequencing has several advantages, first and foremost that it provides sufficient information in a small sequence, enabling cheap profiling of the microbiome (Janda et al., 2007; Cook et al., 2003). Its main disadvantage is the lack of annotation for most microbes, partially due to the lack of whole genomes, and its lack of resolution when it comes to species and strain level differences. Metagenomics sequencing approaches aim to overcome this problem by sequencing all microbial DNA. However, it remains a very expensive method and metagenomics datasets consist of a complex mixture of thousands of microbial genomes, making data analysis challenging. In this thesis, I show that faecal metabolic profiles largely represent the microbial flora and might therefore be used as a novel method to profile gut microbial communities in future studies (for details see chapter 7).

Although twin studies have found a modest genetic influence on some phyla, most of the variation is environmental (Goodrich et al., 2014; Goodrich et al., 2016). The composition of the microbe flora varies a lot across individuals (Zhu et al., 2010; Turnbaugh et al., 2007) and even

between different parts of the body (Kong, 2011). It has a huge influence on many biological processes such as immune response, metabolism, and disease (Zhu et al., 2010; Grice et al., 2012). While the microbiome seems to be relatively stable during adulthood, it changes significantly in later life (Claesson et al., 2011; Biagi et al., 2010). Biagi et al. (2010) observed drastic changes in the gut microbiome of centenarians, namely a general loss of diversity and increased abundance of *bacilli* and *proteobacteria*. The latter were reported to promote inflammation under certain conditions (Round et al., 2009). Similar findings were revealed in other elderly populations, in studies that also considered the dietary and residential situation of elderly patients (Claesson et al., 2012). More recently, the gut microbiome was also found to be associated with frailty (Jackson et al., 2016a).

There is also some preliminary evidence that the gut microbiome associates with renal function (Barrios et al., 2015; Vaziri et al., 2013; Ramezani et al., 2014), however, these findings should be systematically confirmed in future studies.

1.2.8 Phenomics

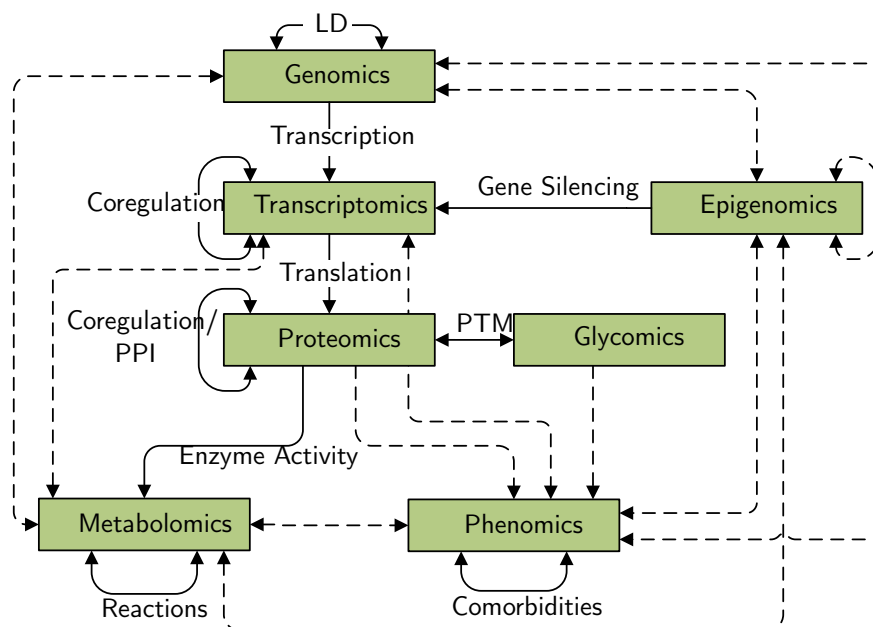
Simultaneously with omics data, the dimension of clinical and lifestyle traits, particularly clinically used intermediate traits, keeps increasing. Epidemiological studies collect thousands of phenotypes beyond omics data types ranging from anthropometric measures to health and lifestyle questionnaires (Moayyeri et al., 2013). Collecting high-dimensional clinical data is crucial to unveiling pleiotropy of genes and interactions among clinical phenotypes such as comorbidities (Houle et al., 2010). Driven by high-throughput technologies, statistical and bioinformatics methods for analysing high dimensional data are becoming available. These facilitate the investigation of many clinical phenotypes in parallel, thus defining the new field of phenomics (Houle et al., 2010).

Phenomics is particularly important for ageing research, as dozens of clinical phenotypes, such as CKD, obesity and blood pressure (Mungreiphy et al., 2011), as well as lifestyle parameters, such as nutrition, smoking and physical activity, are strongly associated with age and amongst each other. Composite measures such as frailty indices combine several of those clinical traits to form a more homogenous phenotype – frailty – from its diverse appearance, which can be considered as a measure for biological age (Mitnitski et al., 2013) (see section 1.1.2.2). However, only multi-variate analyses integrating multiple phenotypes can help to unveil interdependencies among them, such as disease comorbidities (see chapter 9).

1.3 Systems biology

Most of the studies summarised above concentrated on the association of age or age-related diseases with one type of omics data, analysing each of its compounds separately. However, there are strong interdependencies within and between the different omics data (Figure 1.4) and correlations can be observed between practically all levels of biological organisation.

Figure 1.4 Interdependencies of 'omics' layers. Interdependencies can be observed within and between almost all omics data sets. Solid lines indicate biological processes which cause dependencies, while dashed lines represent observed associations. (Picture adapted from Zierer et al. (2015))



Following the central dogma of molecular biology, genomics, transcriptomics and proteomics are correlated 'by definition'. Furthermore, metabolite concentrations are influenced by genetic variants (Shin et al., 2014; Long et al., 2017a) and epigenetic factors (Petersen et al., 2014) mediated by changes in gene expression or enzyme activity. Methylation levels do not only influence the gene expression (Jaenisch et al., 2003), but are also correlated with genetic variants (Bell et al., 2012) and environmental factors (Breitling et al., 2011). My group has recently demonstrated that even the microbe composition is partly under host genetic influence (Goodrich et al., 2014).

Correlations, however, do not only occur between but also within each type of data. For instance, in genomics linkage disequilibrium, which is the correlated occurrence of SNPs, is a ubiquitous phenomenon. Transcription factors often co-regulate the expression of multiple genes (Allocco et al., 2004), and methylation patterns of neighbouring CpG sites were reported to be correlated (Bell et al., 2012). Metabolites are linked by a network of biochemical reactions, causing strong correlations between them (Kruksiek et al., 2011). Even phenotypes often cluster: Comorbidities were shown to affect many diseases, possibly due to shared underlying mechanisms (Goh et al., 2007).

These biological correlations can confound association studies. Consequently, studies regularly identify a plethora of compounds significantly associated with a phenotype of interest. However, the biologically interesting, causal associations are often obscured by this abundance of results. For instance, 153 metabolites were found to be associated with age, but subsequent analyses showed that only 22 of them are independently associated with age (Menni et al., 2013b). Similarly, 21 of 24 measured IgG glycans were correlated with age but only 3 glycans

were associated independently, together explaining 58 % of the variance (Krstic et al., 2014). Analogous results were found for epigenetic data (Weidner et al., 2014).

In contrast to the reductionist approaches, which analyse each component of a high-dimensional dataset separately, systems biology thus aims to analyse ideally all components of a biological process simultaneously, taking their interactions and their intrinsic hierarchical structure into account (Barabási et al., 2004; Ideker et al., 2001). By explicitly analysing interdependencies, systems biology addresses the problem of spurious correlations and aims to identify the underlying causal network, rather than separate components (Cassman, 2005).

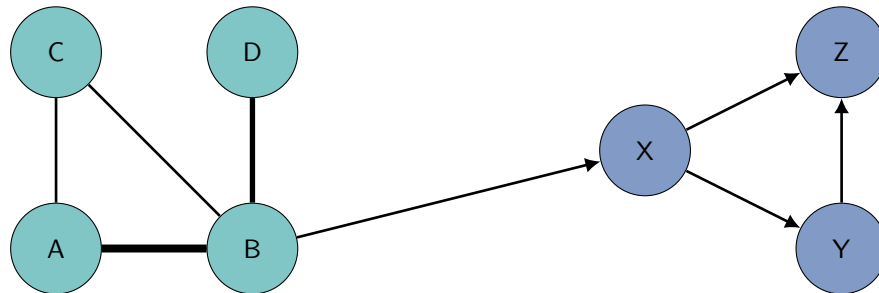
While deducing causality from observational data and even more the holistic modelling of a single eukaryote cell remain far-fetched goals, the field of systems biology is advancing and a multitude of methods is readily available to facilitate integrative data analysis. In the following, I will introduce key concepts of systems biology and their application in ageing research. Methodological details of the methods of interest are further described in section 3.4.

1.3.1 Introduction to graphs

The defining feature of systems biology is the analysis of dependency structures amongst many variables simultaneously. Graphs – or networks – provide the theoretical framework to modelling these interdependencies. A graph $G(V, E)$ is a set of nodes – or vertices V – that represent compounds or actors in the model. Nodes are connected by a set of edges E that describe their dependencies. Both nodes and edges can have different manifestations, which change the meaning and the interpretation of the network and its properties (Figure 1.5). For instance, metabolites interact in chemical reactions, forming a network in which nodes describe the metabolic compounds and edges indicate chemical reactions. In contrast, a gene-regulatory network (GRN) consists of nodes that represent genes and edges that represent the regulation of one gene by another, thus modelling gene expression. Protein-protein interaction (PPI) networks model physical interactions of proteins, such as the formation of dimers or larger protein complexes. Each node and edge can have properties that further describe their behaviour in the model. An edge might have a weight, which for instance indicates the strength of the regulation or a reaction rate, thus forming a weighted network. Edges also might have a direction, indicating causality of the relationship, forming a directed network.

Graphs offer the advantage of accessible visualisation of complex relationships, though their main benefit is the amount of theoretical research addressing the analysis of graphs, providing many methods for analysing graph topology. One common problem is, however, the identification of modules, i.e. densely connected subgraphs. In biological networks, modules correspond to functional units, such as the glycolysis pathway in the metabolic network. Different modules are usually sparsely connected with each other and together form a hierarchical structure in which the distribution of node degrees – the number of edges per node – follows a power-law (Barabási et al., 2004). Hence, most nodes have only few connections while few nodes have many connections. These highly connected nodes are called hubs (Albert et al., 2000; Jeong et al.,

Figure 1.5 Graphs. This is an exemplary graph with two groups of nodes: (1) nodes A – D, and (2) nodes X – Z. For illustration purposes edges in the first module are undirected, while edges in the second module are directed (indicated by arrows). Nodes represent entities in the system while edges represent their relationships. These might have properties that are included in the model, such as weights (indicated by width of the connection).



2001). Hubs can be thought as the main driving factors in a network, that have the potential to quickly affect large portions of the system (see Figure 3.11 for details).

In the following I will discuss common approaches to integrative data analysis using graphs with a particular focus on their application in ageing research.

1.3.2 Analysis of pre-defined graphs

A popular approach to put the results of an association study in a systems biology context is projecting the variables of interest – such as age-related genes, proteins, or metabolites – onto reference networks. The neighbourhood of these target variables and their topological properties can then be assessed using experimentally pre-defined PPIs, GRNs, or metabolic networks. Instead of interpreting individual entities separately, a priori knowledge about gene interactions or metabolic reactions and their grouping according to functions and localization can be used to identify functional network modules that show significant changes in the condition of interest.

Several databases offer collections of experimentally identified interactions that can be used as pre-defined reference networks for enrichment and topology analyses after mapping single entities of interest. In case of PPIs the Human Protein Reference Database provides more than 40,000 (Keshava Prasad et al., 2009), the Database of Interacting Proteins more than 7000 (Xenarios et al., 2002), and the Munich Information centre for Protein Sequences (MIPS) mammalian protein-protein database roughly 1000 hand-curated interactions of human proteins (Pagel et al., 2005). The STRING database combines PPIs from various sources, ranging from experimentally confirmed interactions to text-mining approaches (Franceschini et al., 2013). It currently contains more than 184 million interactions for more than 2000 organisms. Experimentally derived GRNs are provided by the ChIPBase (Yang et al., 2013), which contains six million transcription factor binding sites from >300 experiments. Metabolic reactions are amongst others provided by Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2012), BioCyc (Caspi et al., 2014) and Recon (Thiele et al., 2013).

1.3.2.1 Pathway enrichment analysis

Enrichment analysis is a convenient way to incorporate existing knowledge from biological reference networks without directly analysing the graph topology. Thereby, pre-defined (functional) modules within the reference networks are used to test over-representation of the interesting genes, proteins, or metabolites in these groups. Pathway enrichment follows the idea that certain pre-defined sets of variables, for example genes that are known to be co-regulated or metabolites that are produced by the same biochemical pathways, should be jointly affected by the condition of interest. Consequently, pathway analysis tries to first, identify patterns within the set of results of an association study and second, identify novel pathways of interest, even if a study is underpowered for identifying significant associations of the phenotype with individual variables.

In ageing research, enrichment analysis unveiled an over-expression of genes involved in immune response, lysosome and glycoproteins and an under-expression of mitochondrial and oxidative phosphorylation-related genes in old people compared to young (Magalhães et al., 2009). In human brain tissue, oxidative stress, DNA repair, and inflammation-related genes were shown to be enriched in the set of differentially expressed genes between young and old individuals (Lu et al., 2004).

1.3.2.2 Topological analysis

To avoid pre-defined modules, which do not consider condition-specific interactions, and to enable a more detailed network analysis, the variables of interest can also be mapped directly onto the known PPIs, GRNs, or metabolic networks. Instead of investigating enrichment of the individual variables in pre-defined modules as described before, modules can be identified dynamically based on the measured data. Moreover, additional topological properties of the nodes of interest can be assessed.

Studying human PPI networks revealed that genes which were associated with ageing by homology have higher node degrees and a higher betweenness centrality compared to other genes (Bell et al., 2009). Furthermore, age-related genes were not spread throughout the interactome, but cluster in few, tightly connected modules. Those modules were enriched in DNA damage repair and stress response genes (Kriete et al., 2011). Using a modified PPI network, Wang et al. (2009) showed a connection of the genetic causes of ageing and disease. These results indicate that ageing does not occur due to random errors but is an organized process. Another PPI based approach to data integration was developed by West et al. (2013). They incorporated epigenomic data by assigning DNA methylation sites to each protein in the graph and then identifying modules of differentially methylated genes/proteins in the resulting network. By doing so they avoided pre-defined gene sets as used by enrichment analyses. They found three differentially methylated modules, which were replicated across several tissues. Two of them contained mainly transcription regulating genes, while the third one contained genes related to stem cell differentiation.

1.3.2.3 Integration with data

A drawback of experimentally derived PPI or GRN is that such methods detect up to 50 % false positives while many true interactions are missed (Marbach et al., 2012; Huang et al., 2009). Even more importantly, those reference networks completely ignore the tempero-spacial properties of the interactions. This restricts results to already observed, possibly inactive interactions.

One method to overcome the static nature of PPI networks are Negative-Positive (NP) networks (Xia et al., 2006). These integrate the pre-defined PPI network with transcriptomics data by restricting it to edges between (anti-) correlated proteins/genes. Thereby only those interactions (i.e. edges) that are active under the observed condition are analysed further. Application of this method to the previously mentioned dataset of brain gene expression identified two anti-correlated modules containing cell proliferation and cell differentiation related proteins (Xue et al., 2007). Two other modules consisting of protein processing and immunity-related genes, respectively, were found to be correlated with the cell proliferation module. A more recent study went one step further and restricted a PPI network to highly expressed genes in different stages of ageing for each sample separately, thus generating a set of dynamic binding networks instead of a single network. Even though the global properties of all those graphs were very similar, the centrality of several genes correlated with age (Faisal et al., 2014).

1.3.3 Graph inference

Despite their successful applications, all approaches presented so far still rely on pre-defined, static networks. These are often noisy, specific to particular tissues, or based on model organisms. To overcome these limitations, inferring networks directly from the measured data is the next step.

1.3.3.1 Weighted gene co-expression network analysis

The most straight-forward method to analyse the interdependencies of many variables is a correlation network (CN), where nodes are variables and edges represent significant correlations between them. The weighted gene co-expression network analysis (WGCNA) algorithm (Zhang et al., 2005a) does exactly this, using genes as nodes and defining edges based on the observed correlations of gene expression values. In contrast to normal CNs it does not rely on a strict correlation cut-off to differentiate between ‘significant’ and ‘non-significant’ edges, but instead uses all edges weighted by the strength of their correlation. While the method was initially created for transcriptomics data, it can be easily applied to other omics datasets.

Conducting WGCNA from the previously mentioned gene expression dataset of 30 human frontal cortex samples at different ages and comparing it with another graph derived from an AD transcriptomics study, revealed a significant overlap between healthy ageing and AD (Miller et al., 2008). This suggested a shared molecular basis for both processes, namely synapses-, transport- and transcriptional regulation-related genes.

1.3.3.2 Graphical models

However, when analysing biological datasets CNs suffer from large numbers of spurious correlations due to indirect dependencies between variables (Figure 1.6). For instance, one study found that more than half of all pairwise correlations between 151 metabolites were statistically significant at an alpha level of 1 %, even after conservative correction for multiple testing, in a cohort of >1000 individuals (Krumsiek et al., 2011). Graphical models (GMs) – also known as conditional independence graphs (CIGs) – were proposed to overcome this problem and infer biological meaningful networks in the analysis of metabolomics (Krumsiek et al., 2011; Steuer, 2006) as well as other ‘omics’ data (Fuente et al., 2004; Yuan et al., 2011).

GMs are probabilistic models in which an edge between two variables illustrates their conditional dependence, given all other variables in the model. Implicitly, the absence of an edge represents the conditional independence of the according variables. GMs are well-established in case of binary or multivariate Gaussian distributed data. Ising Models are one instance of GMs for binary data and several algorithms to infer binary graphical models are publicly available, e.g. as R packages (Guo et al., 2010; Ravikumar et al., 2010; Wainwright et al., 2006; Höfling et al., 2009).

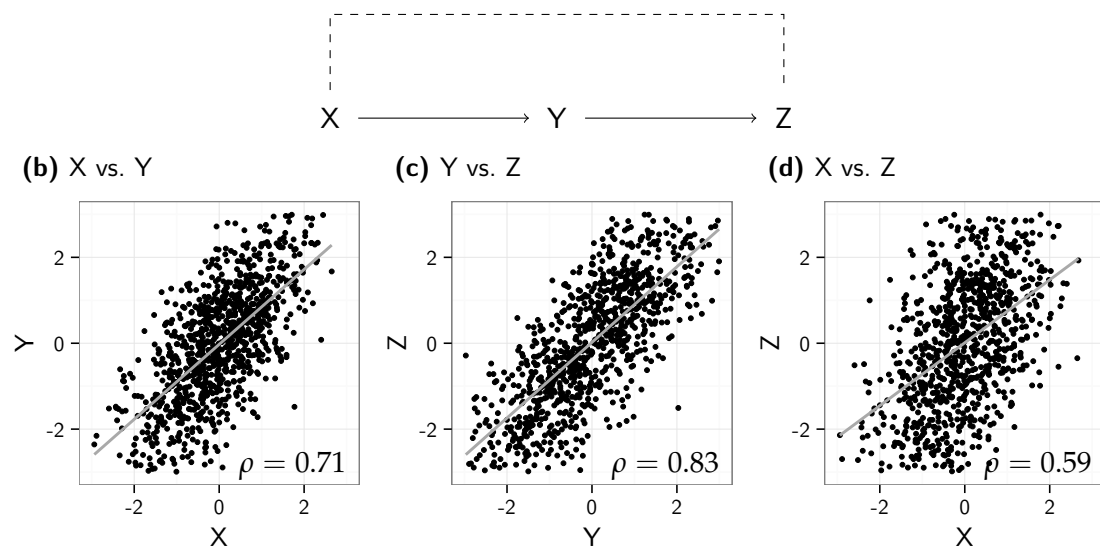
Gaussian graphical models (GGMs) are the most common implementation of CIGs for multivariate Gaussian-distributed data. Edges in a GGM are defined based on partial correlations, i.e. pairwise correlations of variables after adjustment for all other variables in the model. The (weighted) adjacency matrix of a GGM is called precision matrix and can be calculated as the inverse of the covariance matrix of the generating dataset. The absence of an edge is then defined by a partial correlation of 0, i.e. conditional independence of the variables. However, biological data is inherently noisy and empirical partial correlations are rarely exactly zero. More importantly, empirical covariances are often bad estimators of the real covariances when analysing many variables based on few samples ($n \ll p$) and also the covariance matrix is not invertible in these scenarios (Schäfer et al., 2005). Several algorithms were developed to circumvent these problems and infer stable GGMs (Meinshausen et al., 2006; Yuan et al., 2007; Friedman et al., 2008; D’Aspremont et al., 2006; Mazumder et al., 2012). Many of them, such as the well-established graphical lasso (Friedman et al., 2008; Mazumder et al., 2012), use regularization to address the $n \ll p$ problem and avoid over-fitting of graphs. This allows researchers to concentrate on fewer high-confidence interactions (see section 3.4.7).

While GGMs have been used to reconstruct biological networks from metabolomics (Krumsiek et al., 2011) and transcriptomics (Yuan et al., 2011) data, they have not been widely applied in ageing research, yet. One of the few existing studies compared metabolic networks, inferred by GGMs, and demonstrated differences in lipoprotein metabolism between healthy controls and patients with impaired fasting glucose (Valcárcel et al., 2011). These changes often proceed diagnosis and might help to generate hypotheses on the metabolic dysregulation that causes diabetes.

While GMs are well-established for both Gaussian-distributed and binary data, inference of GMs from mixed data, incorporating continuous and categorical variables, is much less common. Shin et al. (2014) addressed this problem of integrating metabolomics and genetic data by first constructing a GGM from metabolite concentrations and then adding SNPs as

Figure 1.6 Spurious correlations. This figure exemplifies the problem of spurious correlations in the analysis of biological data using three random variables, X, Y and Z. X was randomly drawn from a normal distribution, while Y and Z are derived from their respective predecessor by adding normal-distributed, random noise (a). The variables could for instance represent metabolites in a metabolic pathway or genes in a GRN. The correlation X with Y (b) and Y with Z (c), respectively, are causal, whereas the correlation of X with Z (d) is spurious and completely mediated by Y. Normal zero-order correlations, such as Pearson correlation coefficients, cannot detect mediation and, thus, lead to spurious results. In contrast, higher-order correlations consider potential confounding factors and can thus detect such mediation. The higher-order (partial) correlation of X and Z, given Y, is $\rho_{xz|y} = 0.02$ ($p > 0.5$), while the partial correlations $\rho_{xy|z}$ and $\rho_{yz|x}$ remain significant.

(a) Exemplary Reaction System



nodes and connecting them with associated metabolites. The resulting network illustrates the genetic control of the metabolism in an intuitive way. However, it is no longer a GM, and edges do not indicate conditional dependence. Recent developments facilitate the integration of different types of data while maintaining the favourable properties of GGMs, namely mixed graphical models (MGMs) (Lee et al., 2015; Fellinghauer et al., 2013; Chen et al., 2015; Tur et al., 2012). In this thesis, I present the first application of this approach to a multi-omics dataset, analysing comorbidities of age-related diseases (chapter 9).

GGMs as well as MGMs are undirected models and can therefore not be used to infer causal direction. In epidemiological research Mendelian randomization is an increasingly common approach to infer causality from observational data (Brion et al., 2014). It takes advantage of the invariability of gene variants to separate the study population in groups, thus mimicking a randomized controlled trial. Thereby the directionality of an association between two phenotypes can be assessed. Mendelian randomization can be used to further investigate edges of interest,

which might have been identified by graphical models before. However, this approach relies on stable associations with genetic variants and assumes – amongst others – independence of all potential confounding factors. Thus, causal inference from observational data remains challenging.

1.3.3.3 Bayesian networks

Another approach that aims to infer causality from observational data are Bayesian Networks (BNs). Like GMs, BNs are probabilistic models in which edges represent the conditional independence between variables. However, BN are directed acyclic graphs (DAGs) thus distinguishing between an influence of X on Y and the influence of Y on X . In return the acyclicity of the causal graph is an assumption, which might not hold true for biological networks.

The application of BNs on high-throughput transcriptomics data demonstrated the potential of this method to extract meaningful associations from data without prior knowledge (Friedman et al., 2000). Recent developments enable the inference of BNs using genetic variation as instrumental variables, as in Mendelian randomization, in a genome-wide manner to infer directionality of entire network models (Yazdani et al., 2016).

1.3.4 Model biological systems

The ultimate goal of systems biology is not only the qualitative exploration of an organism, but the quantitative modelling of the organism, facilitating *in silico* experiments, hypotheses generation and predictions. Currently, system biological modelling is used only in the field of microorganisms. For instance, Karr et al. (2012) created a model of a mycoplasma genitalium cell, which simulates the cell cycle and predicts metabolite concentrations. However, the model is far from perfect (Freddolino et al., 2012) and too primitive to be adapted to more complex organisms. Nevertheless, some models were developed to investigate ageing phenotypes in very small, well-defined subsystems. McAuley et al. (2009) created a systems biology mark-up language model to investigate the influence of increased cortisol levels on hippocampus activity, which is important in AD. Such models require extensive knowledge about the involved variables and their interactions, which makes their application to complex phenotypes infeasible at the current state of research.

1.4 Conclusion

Due to the increasing age of populations, healthy ageing is a major challenge. The development of high-throughput technologies and growing availability of large ‘omics’ datasets facilitated a multitude of studies investigating the effects of ageing on a molecular level. With the increasing dimensionality and sample sizes of biological datasets, associations studies identified a plethora of compounds associated with ageing and age-related diseases. While these associations do expand our knowledge on diseases, they are often limited by the lack of causal inference.

Systems biology provides tools to analyse high-dimensional datasets integratively, aiming to investigating multi-variate relations between biological compounds. Thereby, multivariate statistics exclude spuriously correlations and the combination of data from various sources permit for stronger hypotheses. Innovative methods, for instance to infer and analyse biological graphs, can help to add structure to the results of an association study, thus facilitating mechanistic insights into disease processes.

CHAPTER 2

Hypothesis, aims, and outline

2.1 Hypothesis

Systems biology approaches can help to gain greater understanding of the causal mechanisms of complex phenotypes, such as biological ageing and related diseases.

2.2 Aims

This hypothesis will be explored through the following aims:

1. Identify associations of ageing and ageing-related diseases, particularly CKD, with molecular, intermediate phenotypes such as glycomics and metabolomics
2. Integrate metabolomics and genetics data with measurements of the gut microbiome to identify microbial metabolites affecting CKD
3. Use systems biology approaches to analyse multi-tissue omics data, particularly metabolomics, to analyse tissue-dependent effects of ageing, and gain insights in the pathology of chronic kidney disease, as example of a highly age-dependent disease
4. Integrate multiple phenotypes with multiple omics data to gain insights in their potentially shared molecular (patho-) mechanisms

2.3 Outline

In this thesis, I aimed to identify metabolic patterns that underlie biological ageing and age-related diseases. First, I analysed metabolites associated with telomere length to identify mechanisms of biological ageing (chapter 4). To further explore the aetiology of CKD, I analysed the relationship of renal function with glycosylation of IgG and lipid metabolism in chapters 5 and 6, respectively. The next step was to analyse the faecal metabolome as a functional marker of the gut microbiome in order to investigate its association with ageing and CKD (chapter 7). In chapter 8, I inferred metabolic networks, integrating metabolomics measurements from plasma, urine and saliva, to identify metabolic processes that differ in renal disease patients compared to the general population. Lastly, in chapter 9, I integrated data from four different omics technologies with phenomics data to investigate comorbidities and underlying (shared) molecular mechanisms of age-related diseases including CKD. Finally, in chapter 10, I summarise the results and implications and discuss its limitations and potential for subsequent studies.

CHAPTER 3

Material and methods

Data from several independent cohorts, measured by different technologies and analysed using various statistical methods were used for each of the chapters of this thesis. In the following, I will describe methods that were used repeatedly throughout the thesis.

3.1 Study populations

In each chapter, different subsets of the available study populations were used, according to data availability. An overview over the cohorts used in each chapter is given in Table 3.1.

3.1.1 TwinsUK

The TwinsUK cohort is a national register of 13,000 adult twins recruited as volunteers without selecting for any particular disease or trait (Moayyeri et al., 2013). Study participants, thus, represent the general British population in terms of lifestyle characteristics such as smoking and nutrition. The study participants are roughly equal numbers of MZ and DZ, predominantly female (82.4 %) same-gender twins. On average, they are 59.5 years old (Table 3.2 and Figure 3.1a) and have a BMI of 26.0 kg/m² (Figure 3.1b). TwinsUK is one of the best characterised cohorts in the world with many phenotypes and cutting edge molecular phenotypes available. Hence, TwinsUK is very well-suited for this project.

Through different projects, including the *Healthy Ageing Twin Study* (Moayyeri et al., 2013) a multitude of age-related phenotypes were collected over a period of more than 20 years. Recruitment is still ongoing and twins are regularly invited to St. Thomas' Hospital, London, UK, for examinations where various clinical tests are performed. Biological samples, including blood, urine, faeces and skin biopsies, are collected during visits. Moreover, questionnaires are regularly sent to all participants and each twin answers on average 1.2 questionnaires per year.

Collected phenotypes include for instance anthropometric measures, such as weight, height and waist and hip circumferences (section 3.2.1). Moreover, many clinical phenotypes were collected: blood pressure was repeatedly measured during hospital visits, lung function was assessed using standard spirometry (see section 3.2.6), and grip strength was assessed using a

Table 3.1 Population characteristics. Different subsets of the available cohorts were used in each of the chapters, according to data availability.
This table gives an overview over all datasets used in each chapter.

Cohort	Chapter 4			Chapter 5			Chapter 6			Chapter 7		Chapter 8		Chapter 9	
	TwinsUK	KORA		TwinsUK	MZ Twins	GenoDiabMar	TwinsUK	KORA	YoungFinns	TwinsUK	SHIP	GANI_MED		TwinsUK	
Sample size, <i>n</i>	3511	904		3212	62	655	1279	1784	2046	786	906	72		510	
MZ:DZ:sing/etons	1654:1360:497			506:1772:934	62:0:0		552:622:105			296:310:180				134:222:154	
Age, years	53.6 ±13.6	60.53±8.78		52.67±14.15	55.45±12.20	69.70±9.32	65.16±8.02	60.87±8.90	41.88±5.00	65.18±7.57	49.94±13.64	64.40±15.38		59.02±9.40	
Female, <i>n</i> (%)	100 %	100 %		94.9 %	96.7 %	39.1 %	95.7 %	51.3 %	54.5 %	95.4 %	56.5 %	30.6 %		100 %	
BMI, kg/m ²	26.21±5.14	27.87±5.25		25.95±4.65	25.64±5.65	30.32±5.05	26.33±4.78	28.15±4.79	26.54±5.05	26.14±4.70	27.35±4.50	30.17±6.79		26.63±4.87	

Table 3.2 Population characteristics of the TwinsUK cohort. Shown are the population characteristics of the entire TwinsUK cohort. Phenotypes as well as ‘omics’ measurements were measured for different subsets of the population at different points in time. Thus, population characteristics differ from study to study.

N	13392
MZ pairs	366 (54.7 %)
DZ pairs	303 (45.3 %)
MZ/DZ	1.2
Females	11,04 (82.4 %)
Age, years	59. (± 16.2)
BMI, kg/m²	26. (± 5.0)

hydraulic hand dynamometer. Blood samples were additionally collected to measure standard clinical biochemistry, such as blood glucose levels, blood cholesterol levels, creatinine levels in serum, which is a marker of renal function (section 3.2.3), and concentrations of *gamma-glutamyltransferase* (GGT) and *alanine aminotransferase* (ALT), which are markers for liver function (section 3.2.5). Moreover, dual-energy X-ray absorptiometry (DXA) scans were performed for 5100 individuals to determine body composition variables such as fat and bone mass (section 3.2.2).

Additionally, several cutting edge molecular phenotypes were also collected: 5654 individuals were genotyped using DNA chips and whole genome sequences were obtained for 1959 twins at an average coverage of $30 \times$ (Long et al., 2017a). Transcriptomics data was measured using the Illumina Human HT-12 V3 Bead chip (Illumina Inc., San Diego, USA) from blood, skin and fat biopsies of 856 twins (Grundberg et al., 2012). Epigenetics measurements were performed using the HumanMethylation450K BeadChip (Illumina Inc., San Diego, USA) from the same samples (Grundberg et al., 2013). Metabolic profiling of blood samples was conducted on several platforms for different fluids (see section 3.3.1 for details). More recently, the gut microbiome was profiled for 2766 twins using 16S sequencing (section 3.3.3) and meta-genomics sequencing was performed for an additional 250 twins (Xie et al., 2016).

This deeply phenotyped resource with multi-omics available allows us to put results of association studies, including GWASs, in a broader context and compare associations of molecular phenotypes with different diseases. More importantly, it enables the combination of multiple omics datasets, including phenomics data, to assess interdependencies between the different omics layers and investigate their role in disease comorbidities. Also, the twin structure of the data enables the estimation the proportion of variance of phenotypes that can be attributed to genetics, by comparing MZ and DZ twins (see section 3.4.1). Thus, TwinsUK was used as discovery cohort for most of the studies in this thesis.

All study participants provided written informed consent. The study has been approved by the local St. Thomas’ Hospital Research Ethics Committee and was carried out in accordance with the approved guidelines.

3.1.2 KORA

To replicate findings, I additionally analysed data from the Cooperative Health Research in the Region of Augsburg (KORA) cohort. The KORA study was established in 1996 as expansion of the MONitoring trends and determinants in CARdiovascular disease (MONICA) project and more than 18,000 participants were recruited since then in different batches (Holle et al., 2005; Wichmann et al., 2005). In 1999, 4261 individuals were newly recruited for the most recent KORA cohort, KORA S4, independently of previous participants. 3080 of these S4 participants (72.3 %) were recalled between 2006 and 2008 (Holle et al., 2005) for the KORA F4 study and 2279 were recalled a second time in 2014 for the KORA FF4 study.

The aim of KORA is to investigate genetic and environmental factors that influence chronic diseases to inform policy and the general public about potential risk factors, with a main focus on cardiovascular diseases, the development of diabetes mellitus, and other metabolic diseases (Holle et al., 2005).

Baseline participants were randomly selected from the city of Augsburg, Germany, and two adjacent counties, thus, representing a mixed urban and rural population whose demographic and socio-economic characteristics roughly reflect those of the average central European population.

All participants were invited to the KORA Study Centre, Augsburg, Germany, where they underwent a standardised interview and a comprehensive medical examination. Similarly to TwinsUK, extensive information on environmental factors, such as smoking and nutrition, and clinical phenotypes, such as renal function and lung function, were collected in each of the KORA cohorts and their follow-ups. Additionally, blood sampling was conducted by the KORA study staff and various ‘omics’ measurements were performed. Genetic variation was measured for all 3080 individuals of the KORA S4 study using several different arrays (Affymetrix, Illumina, Metabochip, Immunochip, Exomechip). Metabolic profiling of blood samples was conducted using various platforms as well, including the platforms provided by Metabolon Inc. (Durham, NC, USA), and Brainshake Ltd. (Vantaa, Finland) (see section 3.3.1 for details).

All study methods were approved by the ethics committee of the Bavarian Chamber of Physicians, Munich, Germany and written informed consent was provided by all participants.

3.1.2.1 GenoDiabMar

GenoDiabMar includes 700 adult T2D patients, who were recruited in the Hospital del Mar (Litoral-Mar area, Barcelona, Spain) between 2012 and 2015, to investigate microvascular complications of diabetes. All study participants are older than 45 years, have been diagnosed with T2D at least 10 years prior to recruitment, and are under anti-diabetic drug treatment. The aim of the study is to research microvascular complications of diabetes. Renal ultrasound, fundoscopy, and measurements of proteinuria were performed for all participants.

3.1.3 SHIP and GANI_MED

The Greifswald Approach to Individualized Medicine (GANI_MED) study consists of several disease cohorts including cardiovascular disease (CVD), T2D and CKD patients. I analysed CKD patients from the GANI_MED renal cohort. Patients were recruited in the *Kuratorium für Dialyse und Nierentransplantation* (Centre for dialysis and renal transplants) and the Department of Internal Medicine at the University of Greifswald. About half of the patients were receiving dialysis at the time of recruitment.

The Study of Health in Pomerania (SHIP) study was – amongst others – designed to be a control cohort for the GANI_MED studies. It is a population cohort based at the University of Greifswald in north-east Germany and was started specifically to research complex diseases and their comorbidities. Study participants were randomly selected from the region of West Pomerania, Germany, stratifying for age, sex, and the place of residence.

All study methods were approved by the ethics committee of the medical faculty of the University of Greifswald and informed written consent was obtained from study participants.

3.1.4 YoungFinns

I additionally analysed data from 2046 individuals from the YoungFinns cohort. The ‘Cardiovascular Risk in Young Finns’ study was started as a pilot in 1978 and the first baseline, incorporating 3596 children aged 3 to 18 years was recruited in 1980. After that follow-ups were conducted approximately every three years (Raitakari et al., 2008).

3.2 Phenotypic measures

Both TwinsUK and KORA are very well phenotyped cohorts with a large variety of measures available. In the following, I will outline phenotypes that are central to this thesis.

3.2.1 Anthropometry

Anthropometric measures were taken by trained study staff during the participant’s visits. Body height and weight as well as hip and waist circumferences were recorded. BMI was calculated as mass in kg over the square of the height in m and is expressed as kg/m².

3.2.2 Body composition

Additionally to classical anthropometric measures, body composition was analysed in depth using dual-energy X-ray absorptiometry (DXA) fan-beam technology (Hologic QDR, Hologic, Inc., Waltham, MA, USA). Participants were asked to wear a gown only, remove all jewellery, and were then placed on the DXA machine in a standardised way. Following the manufacturer’s

guidelines recalibration of the machine using a spine phantom was performed daily and additional calibration with a step phantom was done on a weekly basis. DXA scans were analysed using QDR System Software (version 12.6).

DXA scans were used to quantify bone mineral content (BMC) and bone mineral density (BMD) at different body sites. Additionally, the distribution of fat and lean tissues was computed from the scans (Salamone et al., 2000; Haarbo et al., 1991), including visceral fat mass (Kaul et al., 2012). Due to its particular importance, e.g. as risk factor for the metabolic syndrome (Carr et al., 2004) and mortality (Kuk et al., 2006), visceral fat measurements were additionally validated by comparing DXA-based measurements with computed tomography (CT)-based measurements, which are considered the gold-standard. In a comparison of 63 participants of the TwinsUK cohort both measures correlated 83 %, proving DXA is a cost-effective way to quantify visceral fat (Menni et al., 2016).

Regions-of-interest were defined manually following the manufacturer's guidelines by the same operator. For quantification of visceral fat, the lower horizontal margins were placed above the pelvis, just over the iliac crest. The upper horizontal margins were placed at the half of the distance between the acromions and the iliac crest and the vertical margins were adjusted at the external borders of the body to include all the soft tissue (Menni et al., 2016).

3.2.3 Renal function

Renal function was assessed using the estimated glomerular filtration rate (eGFR), which was calculated from serum creatinine levels (S_{cr}) using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (Levey et al., 2009). Creatinine accumulation indicates decline of filtration in the kidney, however expected levels of creatinine in serum differ between sexes, ethnicities and most importantly age. All these factors are considered in the CKD-EPI equation (equation 3.1), which provides a comparable estimate of the glomerular filtration rate

if black:

$$eGFR = \begin{cases} 166 \times \left(\frac{S_{cr}}{0.7}\right)^{-0.329} \times 0.993^{\text{age}} & \text{if } \text{♀} \wedge S_{cr} \leq 0.7 \text{ mg/dL} \\ 166 \times \left(\frac{S_{cr}}{0.7}\right)^{-1.209} \times 0.993^{\text{age}} & \text{if } \text{♀} \wedge S_{cr} > 0.7 \text{ mg/dL} \\ 163 \times \left(\frac{S_{cr}}{0.9}\right)^{-0.411} \times 0.993^{\text{age}} & \text{if } \text{♂} \wedge S_{cr} \leq 0.9 \text{ mg/dL} \\ 163 \times \left(\frac{S_{cr}}{0.9}\right)^{-1.209} \times 0.993^{\text{age}} & \text{if } \text{♂} \wedge S_{cr} > 0.9 \text{ mg/dL} \end{cases} \quad (3.1)$$

other ethnicities:

$$eGFR = \begin{cases} 144 \times \left(\frac{S_{cr}}{0.7}\right)^{-0.329} \times 0.993^{\text{age}} & \text{if } \varphi \wedge S_{cr} \leq 0.7 \text{ mg/dL} \\ 144 \times \left(\frac{S_{cr}}{0.7}\right)^{-1.209} \times 0.993^{\text{age}} & \text{if } \varphi \wedge S_{cr} > 0.7 \text{ mg/dL} \\ 141 \times \left(\frac{S_{cr}}{0.9}\right)^{-0.411} \times 0.993^{\text{age}} & \text{if } \sigma \wedge S_{cr} \leq 0.9 \text{ mg/dL} \\ 141 \times \left(\frac{S_{cr}}{0.9}\right)^{-1.209} \times 0.993^{\text{age}} & \text{if } \sigma \wedge S_{cr} > 0.9 \text{ mg/dL} \end{cases}$$

In contrast to the previously used Modification of Diet in Renal Disease (MDRD) equation (Levey et al., 1999), the CKD-EPI equation uses two different slopes for the relationship between eGFR and serum creatinine depending on the creatinine concentration and the gender. The second slope was introduced to avoid the underestimation of high eGFR values by the MDRD formula. The CKD-EPI equation is, thus, more suitable to estimate renal function, particularly in the general population with predominantly good renal function (Murata et al., 2011).

Creatinine concentration was measured from serum samples with an enzymatic assay using the Roche modular system (Hoffman-La Roche, Basel, Switzerland) in the pathology department of St. Thomas' Hospital. Creatinine is enzymatically converted to creatine by creatininase, then further to sarcosine by creatinase, which is further degraded to glycine, formaldehyde and hydrogen peroxide by sarcosine oxidase. Peroxidase then catalyses the reaction of the latter to form a green dye and colour intensity, which is directly proportional to creatinine concentration, is photometrically measured.

3.2.4 Leukocyte telomere length

. LTL was measured by quantitative polymerase chain reaction (qPCR) and expressed as ratio of telomere repeat length (T) to a copy number of a single copy gene (S) (Codd et al., 2010; Codd et al., 2013; Cawthon, 2002). QPCR was performed using 25 μ L reactions on a CAS-1200 liquid handling system (Qiagen, United Kingdom) and analysed on a Rotorgene-Q Real Time Thermal Cycler (Qiagen, United Kingdom) for both telomeres (T) and *36B4* as single-copy gene (S). Additionally, a calibrator sample and a no template control were analysed along with the samples. DNA was quantified relative to the calibrator sample to standardise results across plates and telomere length was measured as relative concentration of T over S . Coefficients of variations were estimated at 1.9 % (T), 1.6 % (S), and 2.9 % (T/S ratio), respectively, from duplicate samples (Codd et al., 2010).

3.2.5 Liver function

Circulating levels of the enzymes *gamma-glutamyltransferase* (GGT) (Whitfield, 2001) and *alanine aminotransferase* (ALT) were used as markers for liver function. Both are predominantly expressed in liver and their increased abundance in blood are general markers of liver function, which also is highly age-dependent (Kim et al., 2015).

GGT concentration was measured enzymatically and enzyme concentration was determined photometrically using a Kodak Ektachem dry chemistry analyser (Johnson and Johnson Vitros Ektachem machine).

In vivo, ALT catalyses the reversible reaction of α -ketoglutarate and alanine to glutamate and pyruvate (Karmen et al., 1955). The same reaction is used for enzymatic measurements of ALT concentration in vitro. The created pyruvate is reduced to lactate by lactate dehydrogenase, which leads to oxidation of reduced nicotinamide adenine dinucleotide (NADH) to its oxidised form (NAD⁺). Increasing concentration of NAD⁺ can be measured photometrically as decreased absorbance at 340 nm. Measurements were performed using a Kodak Ektachem dry chemistry analyser (Johnson and Johnson Vitros Ektachem machine).

3.2.6 Spirometry

Lung function was assessed by standard spirometry (Miller et al., 2005) using a Vitalograph Spirometer (Vitalograph, Buckingham, United Kingdom). Heavy cloths, such as coats, are removed prior to the test. Participants sit on a stationary chair and are asked to take a deep breath to fill their lungs before they breath out as fast as they can. Two measures were recorded to quantify lung function: The forced vital capacity (FVC) is the maximal amount of air that can be exhaled after a maximal inspiration. The forced expiratory volume in one second (FEV1) is the maximal amount of air that can be forcefully exhaled in the first second. Both are measured in litres of air at body temperature.

3.3 'Omics' datasets

In the following, I will describe the experimental details of the metabolomics (section 3.3.1), glycomics (section 3.3.2), microbiome (section 3.3.3) and whole genome sequencing (section 3.3.4) datasets that were analysed in this thesis.

3.3.1 Metabolomics

The metabolome describes the collection of all low-molecular-weight molecules – metabolites – in a sample, and metabolomics describes the study thereof (see section 1.2.6). Currently there is no single technique than can measure all circulating metabolites, indeed the number of unique metabolite that is unknown. The available methods fall in two general categories: targeted and untargeted metabolomics. While targeted metabolomics measures a pre-defined, usually limited set of metabolites, untargeted metabolomics aims to measure all metabolites in a sample. Untargeted metabolomics is hypothesis-free, thus facilitating the discovery of unknown and unexpected metabolic features. However, untargeted platforms are usually less sensitive than targeted platforms and do not support absolute quantification of compounds (Menni et al., 2017b) (Figure 3.2).

Figure 3.1 Population characteristics of the TwinsUK cohort. Histograms show the distributions of (a) age and (b) BMI at the time of writing this thesis. Dashed lines indicate the mean values.

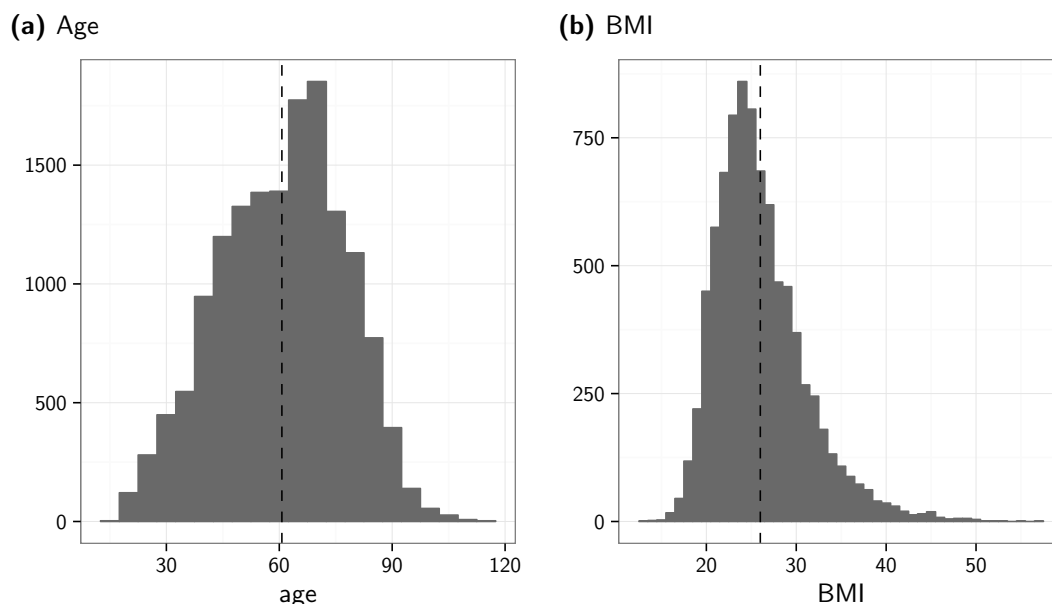
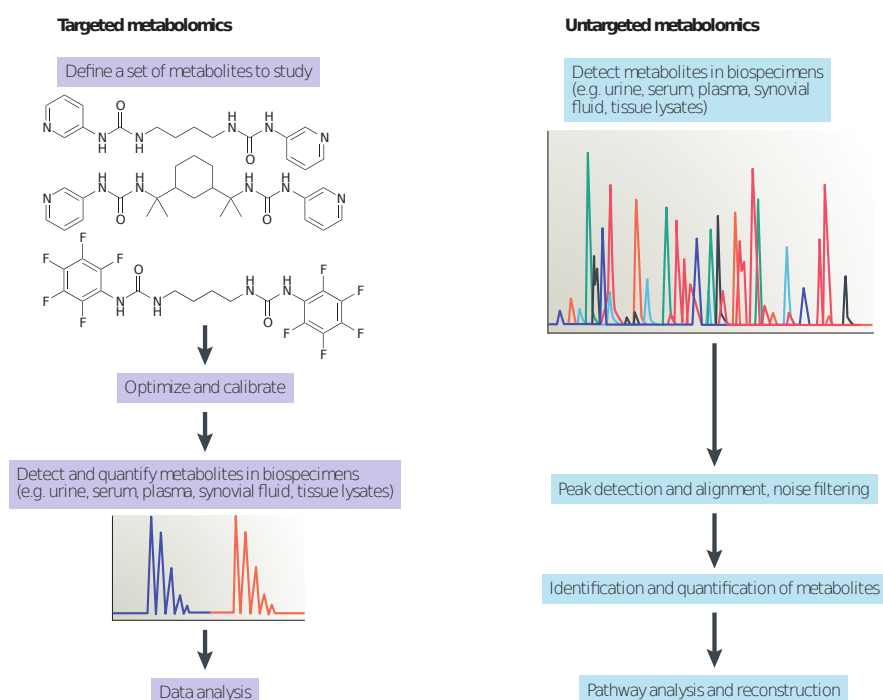


Figure 3.2 Targeted and untargeted metabolomics. Metabolomics approaches can be broadly classified in targeted and untargeted techniques. Targeted platforms aim to measure absolute concentrations of a usually limited set of pre-defined compounds. In contrast, untargeted platforms aim to measure ideally all metabolites in a sample, which comes for the cost of complex compound identification and relative instead of absolute measurements. (Picture previously published in Menni et al. (2017b))



There are two major analytical approaches to metabolomics: nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). NMR exploits the fact that nuclei that are exposed to a magnetic impulse will absorb a portion of the energy and resonate at a specific frequency. The energy emitted due to this resonance can be detected. Depending on the applied magnetic field different nuclei will absorb the radiation energy. In the field of metabolomics usually ^1H nuclei are analysed. Each nucleus within a molecule will absorb – and thus emit – specific amounts of energy with respect to its local environment (which is referred to as *shielding*). The emitted energy can be compared to a standard and is expressed as chemical shift in parts per million (ppm). Consequently, each molecule creates a characteristic pattern of chemical shifts, depending on the configuration of its ^1H nuclei (Figure 3.4). Applying NMR spectroscopy to a biological sample results in a pattern of chemical shifts, from which separate molecules can be identified using bioinformatics methods (Vehtari et al., 2007). NMR measurements do not require elaborate extractions (Markley et al., 2017) and are, thus, very stable (Dumas et al., 2006). However, they lack sensitivity (Markley et al., 2017) and can detect only a small fraction of the human metabolome.

In contrast to NMR-based methods, MS-based methods vary greatly with respect to their exact implementation (Moco et al., 2007). In general, molecules within the sample are ionized, the ions are then sorted according to their mass-to-charge ratio, and finally detected. For each of these steps there exists a multitude of different approaches. Also, MS-based methods rely on extraction of the metabolites from their sample matrices and are usually separated by chromatography prior to injection in the spectrometer. Both steps significantly affect the result of the experiment, for instance by allowing only certain classes of metabolites (e.g. polar or unpolar) to be detected. This additional effort comes with higher sensitivity compared to NMR spectroscopy and allows for detection of diverse sets of metabolites, particularly when combining different approaches (Menni et al., 2017b).

For both analytical approaches, identification of metabolites from the spectra is one of the most challenging steps and one of the major constraints in metabolomics research (Moco et al., 2007). In this thesis, I used two different metabolomics datasets from two commercial providers using different platforms: An untargeted, semi-quantitative, MS-based platform provided by Metabolon Inc. (Durham, NC, USA) and a targeted, quantitative, NMR-based platform provided by Brainshake Ltd. (Vantaa, Finland). Both platforms have been extensively used in research, as they provide identified metabolites rather than the raw, spectral data, thus facilitating straight-forward interpretation of results. In the following, I will describe the analytical procedures of both platforms.

3.3.1.1 Mass-spectrometry (Metabolon)

I analysed multiple datasets that were measured by Metabolon on two different versions of their untargeted platform, which I will refer to as V3 and V4. Both have been extensively used and the experimental protocols have been previously described, e.g. by Shin et al. (2014) and Long et al. (2017a).

Metabolites were extracted by precipitation with methanol under vigorous shaking and subsequent centrifugation. The resulting extract was split in different fractions and analysed on several MS injections. One additional aliquot per sample was kept as backup. For the V3 platform the fractions were analysed in three different injections:

1. Ultra-performance liquid chromatography (UPLC) coupled with a tandem mass spectrometry (MS/MS) on a Thermo Fisher LTQ mass spectrometer (Thermo Fisher, Waltham, MA, USA), using electrospray ionization and a linear ion-trap mass analyser, monitoring for positive ions in acidic extracts
2. The same UPLC-MS/MS platform but instead monitoring for negative ions in basic extracts
3. Gas chromatography (GC)-MS on a Thermo Finnigan Trace DSQ MS (Thermo Finnigan, San Jose, CA, USA) operated at unit mass resolving power with electron impact ionization and a 50 to 750 atomic mass unit scan range

In contrast, the V4 platform uses four MS/MS injections on a Thermo Scientific Q-Exactive high resolution/accurate MS interfaced with a heated electrospray ionization source and Orbitrap mass analyser operated at 35,000 mass resolution. Three of the aliquots are separated using UPLC and the fourth one using hydrophilic interaction chromatography (HILIC) prior to MS analysis.

1. UPLC under acidic positive ion conditions, chromatographically optimized for hydrophilic compounds: The extract was gradient eluted from a Waters Acquity UPLC BEH C18 column using water and methanol, containing 0.05 % perfluoropentanoic acid (PFPA) and 0.1 % formic acid (FA).
2. UPLC under acidic positive ion conditions, chromatographically optimized for hydrophobic compounds: The extract was gradient eluted from the same C18 column using methanol, acetonitrile (ACN), water, 0.05 % PFPA and 0.01 % FA and was operated at an overall higher organic content.
3. UPLC under basic negative ion optimized conditions using a separate dedicated C18 column: Basic extracts were gradient eluted from the column again using methanol and water with 6.5 mM ammonium bicarbonate at pH 8.
4. Negative ionization following elution from a HILIC column: Waters UPLC BEH Amide column using a gradient consisting of water and ACN with 10 mM ammonium formate at a pH of 10.8.

For both versions of the platform metabolites were then identified by comparing the measured features, including retention time and mass-to-charge ratio, against Metabolon's in-house database (Evans et al., 2009; Dehaven et al., 2010). Recurrent features that did not match known metabolites were added as unknown metabolic features, which have the potential to be identified later. Metabolite abundances were measured as area-under-the-curve. These raw area counts were scaled by the run-day median of each metabolite to account for variation resulting from instrument inter-day tuning differences.

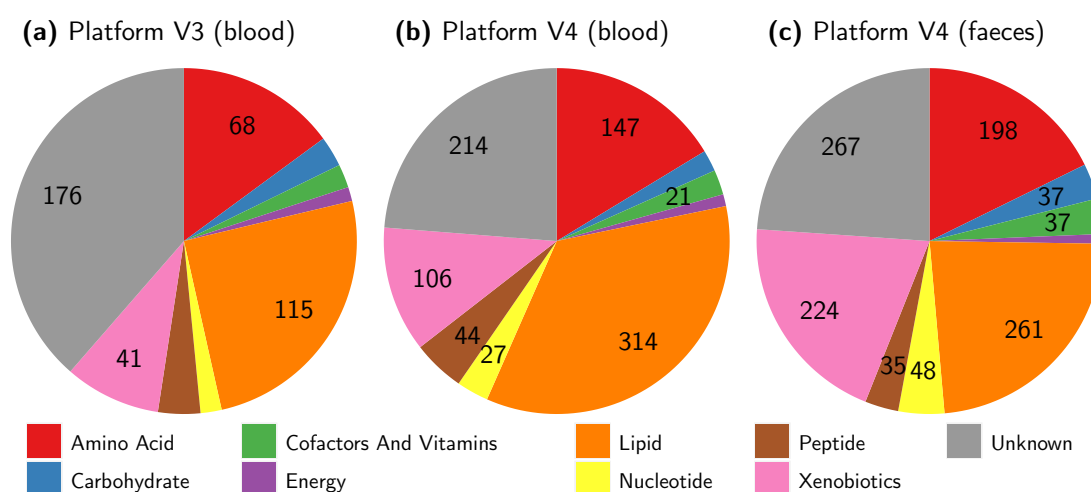
Three types of controls were analysed in concert with the experimental samples: (i) standard samples generated from a pool of human plasma, which has been extensively characterized by Metabolon, were used as technical replicates; (ii) extracted water samples that served as process blanks; and (iii) a cocktail of standards spiked into every analysed sample that enabled instrument performance monitoring. Experimental samples and controls were randomized across the platform run. Instrument variability of the V4 platform was estimated at 5 % from 31 standards. Technical replicates of pooled samples were used to estimate the overall process variability for all metabolites that were detected in at least 90 % of the samples. On average the relative standard deviation (RSD) of these 832 metabolites was 12 % (Table D.1).

Metabolic profiling of fasting blood samples was conducted for 6055 individuals from the TwinsUK cohort using the V3 platform. 5003 of them were fasting plasma samples and 1052 were blood serum samples. The V4 platform was used to profile serum samples of 2069 twins at three different time points. 1971 of the individuals examined using the V4 platform were already examined on the V3 platform. However, the newer V4 platform detects around twice the number of metabolites as the V3 platform, including about 400 additional metabolites with known chemical identity from various chemical pathways (Figure 3.3). Additionally, the V4 platform was used to profile 786 stool samples.

3.3.1.2 NMR-metabolomics (Brainshake)

In contrast to the semi-quantitative measurements provided by Metabolon, Brainshake uses a targeted NMR-platform to provide quantitative measurements of a pre-defined panel of metabolites (Soininen et al., 2015). Like other NMR platforms, the Brainshake platform is less elaborate than MS based metabolomics. Brainshake Ltd. was renamed to Nightingale Health Ltd. during the writing of this thesis.

Figure 3.3 Metabolic pathways covered by Metabolon platforms. While the version V3 of Metabolon's untargeted metabolomics platform detected 280 metabolites with known chemical identity and 176 unknowns in blood samples, version V4 detected 687 known and 214 unknown metabolites. In stool samples the same platform measures 1117 different metabolites, 850 of them with known identity.



Samples were thawed overnight at 4 °C. 100 µL of each sample were mixed with 100 µL of sodium phosphate buffer and subsequently transferred to SampleJet NMR tubes (Bruker, Billerica, MA, USA) using a PerkinElmer JANUS handler (Waltham, MA, USA). Samples were analysed on a Bruker AVANCE III (Bruker, Billerica, MA, USA) 500 MHz NMR spectrometer for 5 min. Two control samples, one plasma sample and one mixture of two low-molecular weight metabolites, were added to each 96 well plate for quality control.

While each NMR spectrum typically consists of several thousand peaks, many of them are redundant and represent the same metabolite. Brainshake uses its proprietary software to derive absolute metabolite concentrations from NMR spectra, independently of external standards (Vehtari et al., 2007). The Bayesian model accounts for correlations among adjacent peaks by summarising them through kernel functions (Figure 3.4). Each metabolite is then calculated as a linear combination of the kernels combined with the overall mean of the spectrum.

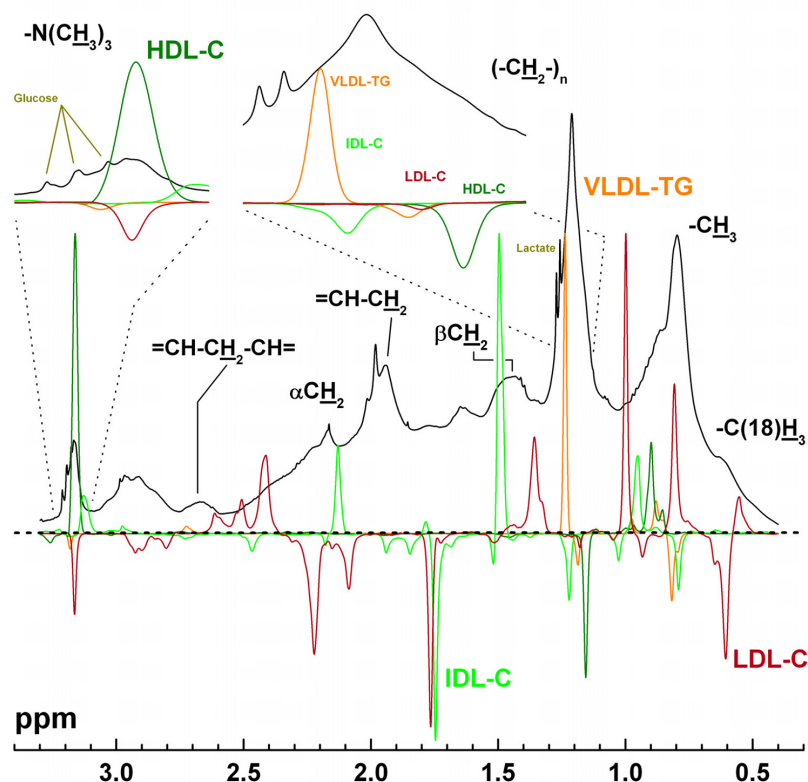
This allows for absolute quantification of a pre-defined set of 144 metabolites, including 98 lipoproteins from 14 lipoprotein subclasses, 9 amino acids and 8 energy related metabolites (Figure 3.5). Moreover, it provides measures of average particle sizes for VLDL, LDL and HDL particles and a semi-quantitative measure of albumin concentration. In addition to these 148 metabolite measures, Brainshake provides 77 metabolite ratios that often represent the fatty acid metabolism better than absolute concentrations (Wurtz et al., 2017; Petersen et al., 2012). These 228 metabolic traits contain most of the metabolites that can be reliably quantified by NMR spectroscopy (Soininen et al., 2015). The platform allows detection of metabolites at concentrations of at least $\approx 10 \mu\text{mol/L}$, depending on the molecular identity. Measurement accuracy of the platform has been shown to be comparable to clinical assays (Wurtz et al., 2017; Vehtari et al., 2007).

3.3.2 Glycosylation of Immunoglobulin G

Glycans are sugar chains that are post-translationally attached to proteins, thus altering their structure and function (section 1.2.5). IgG is a very well-studied glycoprotein in which even subtle changes of the glycans attached to the Fc part can alter its effect between pro- and anti-inflammatory (Gornik et al., 2012). IgG glycans were measured for 4624 twins by Genos Glycoscience Laboratories (Zagreb, Croatia) as previously described (Royle et al., 2008; Kristic et al., 2014).

IgG was isolated using protein G monolithic plates. IgG samples were denatured and glycans were released by incubating samples overnight at 37 °C with 10 µL of 4 % Igepal-CA630 and 1.3 mU of PNGase F in 10 µL 5× phosphate-buffered saline. The released N-glycans were labelled using 25 µL of a freshly prepared labelling mixture of 2-aminobenzamide (2AB) dissolved in a mixture of dimethyl sulfoxide and glacial acetic. An additional 25 µL of reducing agent solution (107.0 mg/mL 2-picoline borane in dimethyl sulfoxide) were added before samples were shaken for 10 min and then incubated for 2 h at 65 °C. 100 µL of each sample were then mixed with 400 µL ACN and free label and reducing agent were removed from the samples using HILIC-solid-phase extraction. Samples were loaded to pre-washed 0.5 µm GHP filter plate, which were then washed seven times using 200 µL ACN/water. Glycans were eluted

Figure 3.4 Exemplary NMR spectrum. In ^1H nuclear magnetic resonance (NMR) spectroscopy samples are exposed to strong magnetic fields to excite protons. These emit resonance energy depending on their local environment. Thus, each molecule produces a characteristic pattern of chemical shifts, which are quantified in parts per million (ppm) (illustrated here in black). Brainshake's proprietary software summarised adjacent peaks using kernels, which are then used to calculate metabolite concentrations. (Picture reproduced with permission from Vehtari et al. (2007))



two times with 100 μL of water. Fluorescence-labelled N-glycans were separated on a Waters ethylene bridged hybrid glycan chromatography column, using a linear gradient of 75 % to 62 % ACN at a flow rate of 0.4 mL/min a 25-minute analytical run at a temperature of 60 $^{\circ}\text{C}$. The system was calibrated using an external standard of hydrolysed and 2AB-labelled glucose oligomers from which the retention times for the individual glycans were converted to glucose units.

Peak integration was performed using an automated pipeline and each chromatogram was manually corrected to assure consistent intervals across samples. The chromatograms were all separated in the same manner into 24 peaks, each of which represents certain glycan structures (Pucic et al., 2011) (Figure 3.7). Additionally, 53 derived traits were calculated as described previously (Pucic et al., 2011) (Table 3.3). These derived traits were defined based on prior knowledge to summarise particular glycosylation features within subgroups of the glycans. They were shown to be more closely related to enzymatic activities and underlying genetic polymorphisms (Lauc et al., 2010).

Figure 3.5 NMR metabolomics provided by Brainshake Ltd.. Brainshake uses NMR spectroscopy to quantify absolute concentrations of 144 metabolites and (filled bullets) along with 80 metabolite ratios (open bullets). (Picture reproduced with permission from Soinenen et al. (2015))

METABOLIC MEASURES

Ketone bodies (mmol/l)

- Acetate
- Acetoacetate
- 3-hydroxybutyrate

Glycolysis related metabolites (mmol/l)

- Glucose
- Lactate
- Pyruvate
- Citrate
- Glycerol

Inflammation (mmol/l)

- Glycoprotein acetyls, mainly α 1-acid glycoprotein

Fatty acids and saturation

- Total fatty acids
- Estimated fatty acid chain length
- Estimated degree of unsaturation

Fatty acids (mmol/l and % of total FAs)

- Omega-3 fatty acids
- Omega-6 fatty acids
- Polyunsaturated fatty acids
- Monounsaturated fatty acids; 16:1, 18:1
- Saturated fatty acids
- Docosahexaenoic acid; 22:6
- Linoleic acid; 18:2
- Conjugated linoleic acid

LMWM

Amino acids (mmol/l)

- Alanine
- Glutamine
- Glycine
- Histidine

LMWM

Branched-chain amino acids

- Isoleucine
- Leucine
- Valine

Aromatic amino acids

- Phenylalanine
- Tyrosine

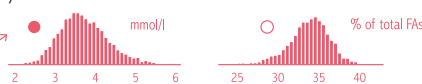
LMWM

Cholesterol (mmol/l)

- VLDL cholesterol
- LDL cholesterol
- HDL cholesterol
- HDL₂ cholesterol
- HDL₃ cholesterol
- Cholesterol
- Free cholesterol
- Esterified cholesterol
- Remnant cholesterol

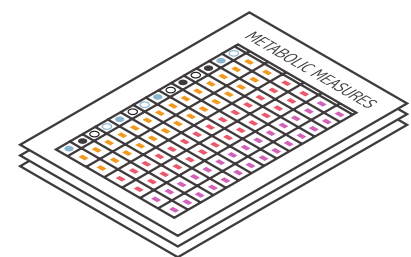
LIPID

LIPID



Apolipoproteins (g/l)

- ApoA-I
- ApoB
- ApoB/ApoA-I



Fluid balance

- Creatinine (mmol/l)
- Albumin (signal area)

LIPID

LMWM

Glycerides & phospholipids (mmol/l)

- VLDL triglycerides
- LDL triglycerides
- HDL triglycerides
- Triglycerides
- Diglycerides
- Phosphoglycerides
- Ratio of diglycerides to triglycerides
- Ratio of triglycerides to phosphoglycerides
- Phosphatidylcholine and other choline
- Sphingomyelins
- Total choline

LIPID

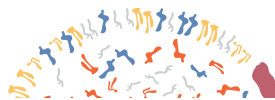
LIPID

Lipoprotein particle size (nm)

- Mean diameter of VLDL particles
- Mean diameter of LDL particles
- Mean diameter of HDL particles

LIPID

14 LIPOPROTEIN SUBCLASSES



12 lipid measures for each subclass

- * Esterified cholesterol (mmol/l and % of total lipids)
- * Free cholesterol (mmol/l and % of total lipids)
- * Triglycerides (mmol/l and % of total lipids)
- * Phospholipids (mmol/l and % of total lipids)

- * Total cholesterol (mmol/l and % of total lipids)
- * Total lipids (mmol/l)
- * Particle concentration (μ mol/l)

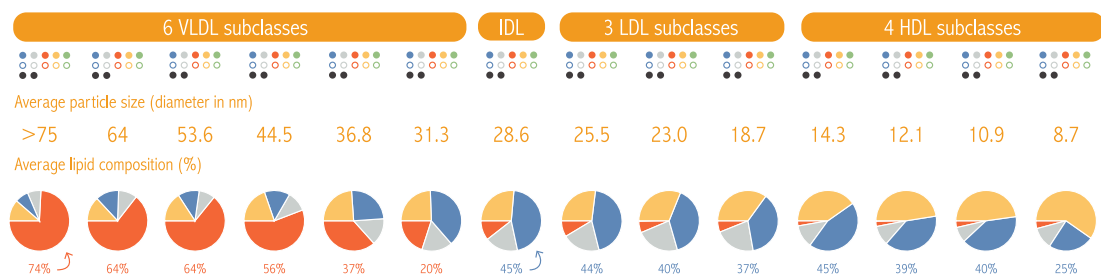


Figure 3.6 The structure of Immunoglobulin G. Structure of *immunoglobulin G* (IgG) with heavy chains in green and light chains in blue. A Fc-bound glycans is represented in yellow (picture created from the PDB structure 1HZH (Saphire, 2001) using Jmol).

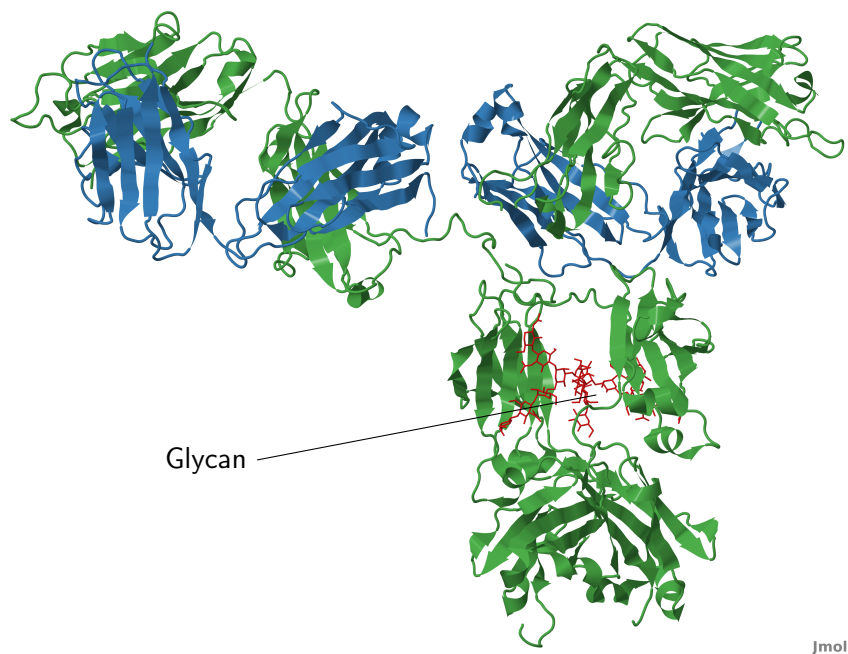


Figure 3.7 Immunoglobulin G glycan chromatogram. Example chromatogram of *immunoglobulin G* (IgG) glycan measurement (published before in Lauc et al. (2013)).

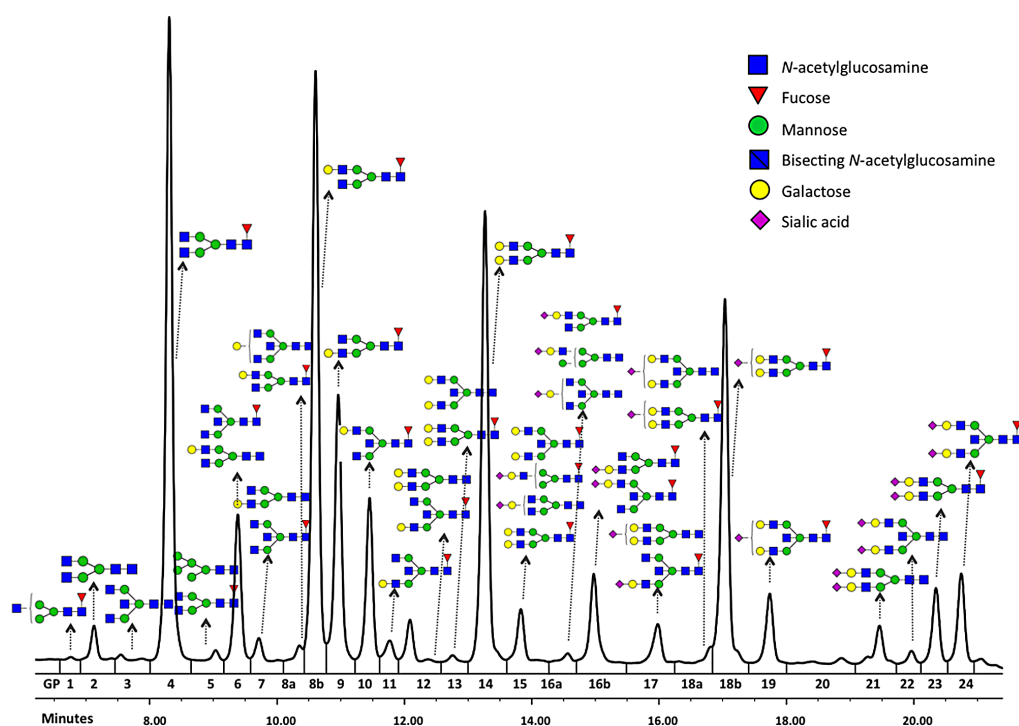


Table 3.3 IgG glycan traits. The table lists all IgG glycan traits that were analysed in this thesis. Traits in (a) are directly measured as area under the curve from the chromatograms (Figure 3.7). All traits in the subsequent tables were derived from the original 24 measures and summarise certain features of glycosylation.

(a) Total Glycans

ID	Trait	Description	Formula
IGP1	GP1	The percentage of FA1 glycan in total IgG glycans	
IGP2	GP2	The percentage of A2 glycan in total IgG glycans	
IGP3	GP4	The percentage of FA2 glycan in total IgG glycans	
IGP4	GP5	The percentage of M5 glycan in total IgG glycans	
IGP5	GP6	The percentage of FA2B glycan in total IgG glycans	
IGP6	GP7	The percentage of A2G1 glycan in total IgG glycans	
IGP7	GP8	The percentage of FA2[6]G1 glycan in total IgG glycans	
IGP8	GP9	The percentage of FA2[3]G1 glycan in total IgG glycans	
IGP9	GP10	The percentage of FA2[6]BG1 glycan in total IgG glycans	
IGP10	GP11	The percentage of FA2[3]BG1 glycan in total IgG glycans	
IGP11	GP12	The percentage of A2G2 glycan in total IgG glycans	
IGP12	GP13	The percentage of A2BG2 glycan in total IgG glycans	
IGP13	GP14	The percentage of FA2G2 glycan in total IgG glycans	
IGP14	GP15	The percentage of FA2BG2 glycan in total IgG glycans	
IGP15	GP16	The percentage of FA2G1S1 glycan in total IgG glycans	
IGP16	GP17	The percentage of A2G2S1 glycan in total IgG glycans	
IGP17	GP18	The percentage of FA2G2S1 glycan in total IgG glycans	
IGP18	GP19	The percentage of FA2BG2S1 glycan in total IgG glycans	
IGP19	GP20	Structure not determined	
IGP20	GP21	The percentage of A2G2S2 glycan in total IgG glycans	
IGP21	GP22	The percentage of A2BG2S2 glycan in total IgG glycans	
IGP22	GP23	The percentage of FA2G2S2 glycan in total IgG glycans	
IGP23	GP24	The percentage of FA2BG2S2 glycan in total IgG glycans	

(b) Total Glycans - Derived Traits

ID	Trait	Description	Formula
IGP24	FGS / (FG + FGS)	The percentage of sialylation of fucosylated galactosylated structures without bisecting GlcNAc in total IgG glycans	$(GP16 + GP18 + GP23) / (GP16 + GP18 + GP23 + GP8 + GP9 + GP14)$
IGP25	FBGS / (FBG + FBGS)	The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans	$(GP19 + GP24) / (GP19 + GP24 + GP10 + GP11 + GP15) * 100$
IGP26	FGS / (F + FG + FGS)	The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans	$(GP16 + GP18 + GP23) / (GP16 + GP18 + GP23 + GP4 + GP8 + GP9 + GP14)$
IGP27	FBGS / (FB + FBG + FBGS)	The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans	$(GP19 + GP24) / (GP19 + GP24 + GP6 + GP10 + GP11 + GP15) * 100$
IGP28	FG1S1 / (FG1 + FG1S1)	The percentage of monosialylation of fucosylated monogalactosylated structures without bisecting GlcNAc in total IgG glycans	$GP16 / (GP16 + GP8 + GP9) * 100$
IGP29	FG2S1 / (FG2 + FG2S1 + FG2S2)	The percentage of monosialylation of fucosylated digalactosylated structures without bisecting GlcNAc in total IgG glycans	$GP18 / (GP18 + GP14 + GP23) * 100$
IGP30	FG2S2 / (FG2 + FG2S1 + FG2S2)	The percentage of disialylation of fucosylated digalactosylated structures without bisecting GlcNAc in total IgG glycans	$GP23 / (GP23 + GP14 + GP18) * 100$
IGP31	FBG2S1 / (FBG2 + FBG2S1 + FBG2S2)	The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans	$GP19 / (GP19 + GP15 + GP24) * 100$
IGP32	FBG2S2 / (FBG2 + FBG2S1 + FBG2S2)	The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans	$GP24 / (GP24 + GP15 + GP19) * 100$
IGP33	$f_{totalS1} / f_{totalS2}$	Ratio of all fucosylated monosialylated and disialylated structures (+/- bisecting GlcNAc) in total IgG glycans	$(GP16 + GP18 + GP19) / (GP23 + GP24)$
IGP34	FS1 / FS2	Ratio of fucosylated monosialylated and disialylated structures (without bisecting GlcNAc) in total IgG glycans	$(GP16 + GP18) / GP23$
IGP35	FBS1 / FBS2	Ratio of fucosylated monosialylated and disialylated structures (with bisecting GlcNAc) in total IgG glycans	$GP19 / GP24$
IGP36	$FBS_{total} / FGS_{total}$	Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc in total IgG glycans	$(GP19 + GP24) / (GP16 + GP18 + GP23)$
IGP37	FBS1 / FS1	Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc in total IgG glycans	$GP19 / (GP16 + GP18)$
IGP38	FBS1 / (FS1 + FBS1)	The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans	$GP19 / (GP16 + GP18 + GP19)$
IGP39	FBS2 / FS2	Ratio of fucosylated disialylated structures with and without bisecting GlcNAc in total IgG glycans	$GP24 / GP23$
IGP40	FBS2 / (FS2 + FBS2)	The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans	$GP24 / (GP23 + GP24)$

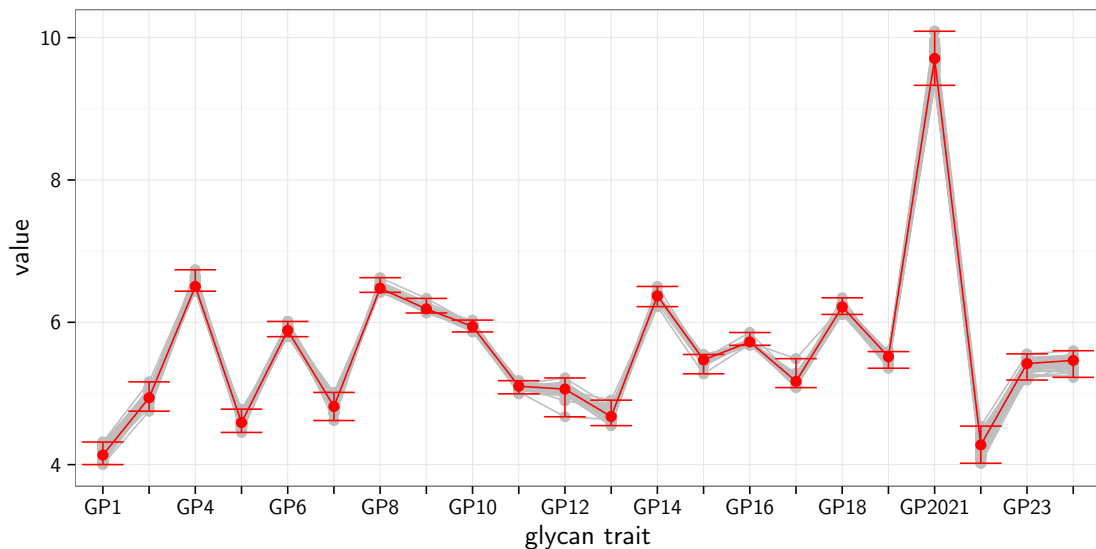
(c) Neutral Glycans

ID	Trait	Description	Formula
IGP41	GP1 ⁿ	The percentage of FA1 glycan in total neutral IgG glycans (GP ⁿ)	$GP1/GP^n \cdot 100$
IGP42	GP2 ⁿ	The percentage of A2 glycan in total neutral IgG glycans (GP ⁿ)	$GP2/GP^n \cdot 100$
IGP43	GP4 ⁿ	The percentage of FA2 glycan in total neutral IgG glycans (GP ⁿ)	$GP4/GP^n \cdot 100$
IGP44	GP5 ⁿ	The percentage of M5 glycan in total neutral IgG glycans (GP ⁿ)	$GP5/GP^n \cdot 100$
IGP45	GP6 ⁿ	The percentage of FA2B glycan in total neutral IgG glycans (GP ⁿ)	$GP6/GP^n \cdot 100$
IGP46	GP7 ⁿ	The percentage of A2G1 glycan in total neutral IgG glycans (GP ⁿ)	$GP7/GP^n \cdot 100$
IGP47	GP8 ⁿ	The percentage of FA2[6]G1 glycan in total neutral IgG glycans (GP ⁿ)	$GP8/GP^n \cdot 100$
IGP48	GP9 ⁿ	The percentage of FA2[3]G1 glycan in total neutral IgG glycans (GP ⁿ)	$GP9/GP^n \cdot 100$
IGP49	GP10 ⁿ	The percentage of FA2[6]BG1 glycan in total neutral IgG glycans (GP ⁿ)	$GP10/GP^n \cdot 100$
IGP50	GP11 ⁿ	The percentage of FA2[3]BG1 glycan in total neutral IgG glycans (GP ⁿ)	$GP11/GP^n \cdot 100$
IGP51	GP12 ⁿ	The percentage of A2G2 glycan in total neutral IgG glycans (GP ⁿ)	$GP12/GP^n \cdot 100$
IGP52	GP13 ⁿ	The percentage of A2BG2 glycan in total neutral IgG glycans (GP ⁿ)	$GP13/GP^n \cdot 100$
IGP53	GP14 ⁿ	The percentage of FA2G2 glycan in total neutral IgG glycans (GP ⁿ)	$GP14/GP^n \cdot 100$
IGP54	GP15 ⁿ	The percentage of FA2BG2 glycan in total neutral IgG glycans (GP ⁿ)	$GP15/GP^n \cdot 100$

(d) Neutral Glycans - Derived Traits

ID	Trait	Description	Formula
IGP55	G0 ⁿ	The percentage of agalactosylated structures in total neutral IgG glycans	$(GP1^n : GP4^n + GP6^n)$
IGP56	G1 ⁿ	The percentage of monogalactosylated structures in total neutral IgG glycans	$(GP7^n : GP11^n)$
IGP57	G2 ⁿ	The percentage of digalactosylated structures in total neutral IgG glycans	$(GP12^n : GP15^n)$
IGP58	F ⁿ total	The percentage of all fucosylated structures (+/- bisecting GlcNAc) in total neutral IgG glycans	$(GP1^n + GP4^n + GP6^n + GP8^n + GP9^n + GP10^n + GP11^n + GP14^n + GP15^n)$
IGP59	FG0 ⁿ total / G0 ⁿ	The percentage of fucosylation of agalactosylated structures in total neutral IgG glycans	$(GP1^n + GP4^n + GP6^n) / G0^n * 100$
IGP60	FG1 ⁿ total / G1 ⁿ	The percentage of fucosylation of monogalactosylated structures in total neutral IgG glycans	$(GP8^n + GP9^n + GP10^n + GP11^n) / G1^n * 100$
IGP61	FG2 ⁿ total / G2 ⁿ	The percentage of fucosylation of digalactosylated structures in total neutral IgG glycans	$(GP14^n + GP15^n) / G2^n * 100$
IGP62	F ⁿ	The percentage of fucosylated structures (without bisecting GlcNAc) in total neutral IgG glycans	$(GP1^n + GP4^n + GP8^n + GP9^n + GP14^n)$
IGP63	FG0 ⁿ / G0 ⁿ	The percentage of fucosylation of agalactosylated structures (without bisecting GlcNAc) in total neutral IgG glycans	$(GP1^n + GP4^n) / G0^n * 100$
IGP64	FG1 ⁿ / G1 ⁿ	The percentage of fucosylation of monogalactosylated structures (without bisecting GlcNAc) in total neutral IgG glycans	$(GP8^n + GP9^n) / G1^n * 100$
IGP65	FG2 ⁿ / G2 ⁿ	The percentage of fucosylation of digalactosylated structures (without bisecting GlcNAc) in total neutral IgG glycans	$GP14^n / G2^n * 100$
IGP66	FB ⁿ	The percentage of fucosylated structures (with bisecting GlcNAc) in total neutral IgG glycans	$(GP6^n + GP10^n + GP11^n + GP15^n)$
IGP67	FBG0 ⁿ / G0 ⁿ	The percentage of fucosylation of agalactosylated structures (with bisecting GlcNAc) in total neutral IgG glycans	$GP6^n / G0^n * 100$
IGP68	FBG1 ⁿ / G1 ⁿ	The percentage of fucosylation of monogalactosylated structures (with bisecting GlcNAc) in total neutral IgG glycans	$(GP10^n + GP11^n) / G1^n * 100$
IGP69	FBG2 ⁿ / G2 ⁿ	The percentage of fucosylation of digalactosylated structures (with bisecting GlcNAc) in total neutral IgG glycans	$GP15^n / G2^n \cdot 100$
IGP70	FB ⁿ / F ⁿ	Ratio of fucosylated structures with and without bisecting GlcNAc in total neutral IgG glycans	$FB^n / F^n \cdot 100$
IGP71	FB ⁿ / F ⁿ total	The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans	$FB^n / F^n \text{ total} \cdot 100$
IGP72	F ⁿ / (B ⁿ + FB ⁿ)	Ratio of fucosylated non-bisecting GlcNAc structures and all structures with bisecting GlcNAc in total neutral IgG glycans	$F^n / (GP13^n + FB^n)$
IGP73	B ⁿ / (F ⁿ + FB ⁿ)	Ratio of structures with bisecting GlcNAc and all fucosylated structures (+/- bisecting GlcNAc) in total neutral IgG glycans	$GP13^n / (F^n + FB^n) * 1000$
IGP74	FBG2 ⁿ / FG2 ⁿ	Ratio of fucosylated digalactosylated structures with and without bisecting GlcNAc in total neutral IgG glycans	$GP15^n / GP14^n$
IGP75	FBG2 ⁿ / (FG2 ⁿ + FBG2 ⁿ)	The incidence of bisecting GlcNAc in all fucosylated digalactosylated structures in total neutral IgG glycans	$GP15^n / (GP14^n + GP15^n) * 100$
IGP76	FG2 ⁿ / (BG2 ⁿ + FBG2 ⁿ)	Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc in total neutral IgG glycans	$GP14^n / (GP13^n + GP15^n)$
IGP77	BG2 ⁿ / (FG2 ⁿ + FBG2 ⁿ)	Ratio of digalactosylated structures with bisecting GlcNAc and all fucosylated digalactosylated structures (+/- bisecting GlcNAc) in total neutral IgG glycans	$GP13^n / (GP14^n + GP15^n) * 1000$

Figure 3.8 Reproducibility of IgG glycan measurements. *Immunoglobulin G* (IgG) glycan measurements are shown for 141 technical replicates of a pooled sample. The average of all samples as well as the minima and maxima, respectively, are shown in red.



The glycan traits were global normalised to account for a multiplicative error. To this end a reference sample was estimated as median sample for each glycan. For each of the measured sample the average deviance from this reference was calculated as median of the fold-changes of each of the 24 measured glycan traits over the reference trait. This average deviance was the used as sample-specific scaling factor for each of the glycan traits. All measurements were corrected for batch effects using the ComBat function implemented in the R-package *sva* (Johnson et al., 2007). Reproducibility of the glycan measurements as well as the validity of this normalisation was assessed using 141 technical replicates across 30 plates from a pooled sample (Figure 3.8). The average RSD of the normalised glycan measurements was estimated at 0.8 % (ranging between 0.3 % to 1.9 %).

3.3.3 Microbial sequencing

Faecal samples were collected to assess the gut microbial composition. Sample collection kits were sent out to participants prior to visits. Participants were asked to keep samples refrigerated in ice packs until arrival at the hospital, where samples were frozen at -80°C (mostly within 24 hours from collection). About 15 % of the study population sent samples by post.

Frozen samples were shipped to Cornell University, where microbial sequencing was performed in the Cornell Biotechnology Resource Center Genomics Facility as previously described (Goodrich et al., 2014; Goodrich et al., 2016; Jackson et al., 2016b). DNA was isolated from the samples using the PowerSoil kit (MO BIO Laboratories, Carlsbad, CA, USA). *515F* and *806R* primers were used to amplify the V4 region of bacterial 16S ribosomal RNA (rRNA) gene by polymerase chain reaction (PCR) (Caporaso et al., 2011). Sample specific barcodes were

added to reads to enable for multiplexed sequencing. The barcoded, pooled samples were sequenced on an Illumina MiSeq machine (Illumina, Inc., San Diego, CA, USA), generating 250 bp paired-end reads. Samples were merged and filtered for a minimum overlap of 200 nt using `fastq join` within QIIME (Caporaso et al., 2010). Then, joined reads were demultiplexed and barcodes were removed, again using QIIME. Reads were subsequently filtered to remove chimeric sequences produced during PCR using USEARCH (Edgar et al., 2011). Samples with less than 10,000 reads were excluded from further analyses; the remaining samples were on average sequenced at a depth of 81,388. All reads that passed QC were grouped to operational taxonomical units (OTUs) using the `Sumac1ust` algorithm, implemented in QIIME. De novo OTUs were shown to cluster reads more accurately at 97 % sequence identity (Westcott et al., 2015), and `Sumac1ust` was among the best performing greedy clustering algorithms in previous comparisons using the TwinsUK data (Jackson et al., 2016b). To assign taxonomy representative sequences of OTUs were aligned against the Greengenes 13_8 database (DeSantis et al., 2006) with a similarity threshold of 97 % using UCLUST (Edgar, 2010).

The de novo clustering across all 3384 samples within the TwinsUK cohort produced $\approx 300,000$ OTUs after singleton removal. However, most of these OTUs were found in very few samples (table density 0.002). 581 OTUs were present in at least 25 % of the samples were included in further analyses (table density 0.547). OTUs counts were converted to relative abundances by dividing read counts over the total number of reads per sample. A pseudo count of 10^{-6} was added prior to log transformation to account for zero counts. The relative abundances were then adjusted for the technical covariates sequencing run, sequencing depth, individual who extracted the DNA, individual who loaded the DNA and sample collection method by fitting linear models with OTU abundances as dependent variable. Residuals of these models were used for further analyses. Additionally, OTUs were collapsed and each taxonomical level and taxonomic abundances were normalised in the same way.

Microbial communities are often described using diversity measures, which describe the within-sample diversity (alpha diversity) and the between-sample diversity (beta-diversity). Diversities were calculated from the complete OTU table and each sample was rarefied to a depth of 10,000 reads 50 times to account for sequencing depth. Diversity metrics were calculated for each sample in each table and the mean across all tables taken as the final measure. Alpha diversity was calculated as Shannon diversity H' , which takes into account frequency and evenness of different species found in a sample (equation 3.2).

$$H' = - \sum_{i=1}^R p_i \ln p_i \quad (3.2)$$

where p_i is the frequency of the i -th (of R) species. Beta diversity was quantified using the unweighted UniFrac algorithm (Lozupone et al., 2005). It assesses co-abundances of species between samples while taking into account the phylogenetic similarity of species.

3.3.4 Whole genome sequencing

Whole genome sequencing was conducted for 2053 individuals from the TwinsUK cohort by HLI Inc. (San Diego, CA, USA) (Long et al., 2017a).

DNA was extracted using the Chemagic DNA Blood400 kit (PerkinElmer chemagen, Baesweiler, Germany), eluted in 50 μ L elution buffer and stored at 4 °C. DNA libraries were prepared using TruSeq Nano DNA HT kit (Illumina Inc., San Diego, CA, USA) following the manufacturer's recommendations, and normalised to 2–3.5 nM. Samples were pooled in groups of six and sequenced on an Illumina HiSeqX sequencer (Illumina Inc., San Diego, USA), producing 150 bp pair-end reads at a coverage of 30 \times .

Reads were aligned to the human genome (build hg38) and variant calling was conducted using ISIS Analysis Software (v. 2.5.26.13; Illumina) (Raczy et al., 2013). Missing variants were assumed to be homozygous for the reference allele (Telenti et al., 2016). The ratio of heterozygous to homozygous variants was calculated from all 'PASS' single nucleotide variants (SNVs), and genomes with a ratio higher than 2.5 were excluded. 1960 individuals of European descent and consistent self-reported and genetic zygosity were kept for further analysis. VCF files of those individuals including all autosomal variants were then merged into a single binary file using bcftools (Li, 2011), whereby multi-allelic variants are split into multiple bi-allelic variants and indels are left normalized. High confidence regions of the genome were defined for the entire cohort. These had to have at least 90 % 'PASS' calls in three sets, each of them consisting of 100 randomly selected genomes. Variants within these regions with a minor allele frequency (MAF) of at least 0.1 % were converted to plink files for further analysis.

3.4 Statistical analysis

Several statistical methods were used repeatedly throughout this thesis. Here I will introduce these methods, while more specific methods will be described in the respective chapters.

3.4.1 Heritability

A commonly asked question, particularly when dealing with novel measurements, is how much of the observed phenotypic variance can be attributed to genetics. The proportion of this genetically caused variance over the total variance is referred to as heritability. There are different methods to estimate the heritability of a trait, many of which take advantage of family structure and particularly twins. Twin modelling utilises the fact that MZ twins are genetically identical and DZ twins share only half of their genetic variation, while both are exposed to the same – or at least very similar – environmental conditions (Figure 3.9). Consequently, the correlation between MZ twins (r_{MZ}) can be attributed to the heritable component of the trait A and the shared environmental component of the trait C , while the correlation between DZ twins (r_{DZ}) is caused by half of the heritable component and the environmental component. The remaining variance is caused by the unique environment of each individual E , which is not shared between twins (Rijsdijk et al., 2002). The correlations between twins are calculated as

$$r_{MZ} = A + C + 2 \times \text{Corr}(A, C)$$

$$r_{DZ} = \frac{1}{2}A + C + 2 \times \text{Corr}(\frac{1}{2}A, C)$$

Assuming that the correlation of genetic and environmental influences $\text{Corr}(A, C)$ is minimal, one can estimate the heritable component A and the common environmental component C from the observed correlations between MZ and DZ twins using Falconer's formula (Falconer, 1960; Rijdsdijk et al., 2002)

$$A = 2 \times (r_{MZ} - r_{DZ})$$

$$C = 2 \times r_{DZ} - r_{DZ}$$

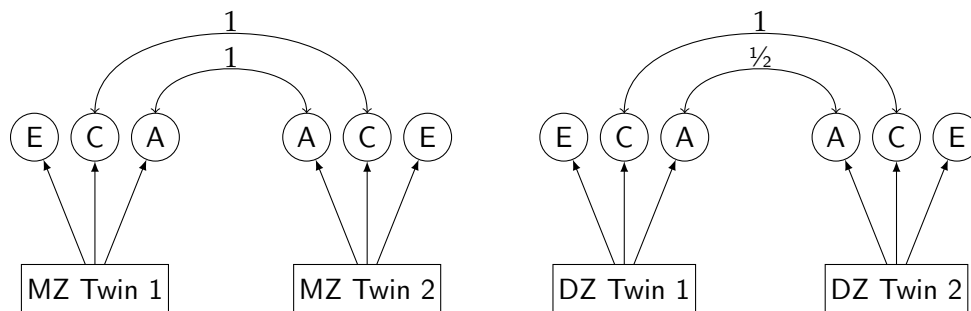
$$E = 1 - A - C$$

Structural equation modelling (SEM) provides a more sophisticated solution for estimating the heritability from twins data. SEM is a very general framework that facilitates the analysis of the correlation structure between variables as well as potential latent factors underlying them (Livote, 2009). In this case, SEM improves the estimation of heritability over Falconer's method by correcting for confounding factors and providing confidence intervals for the point estimates of heritability and common environmental influences (Neale et al., 1997; Martin et al., 1977). By using SEM assumptions about the relationships of A , C and E components between twins must be explicitly formulated (as depicted in Figure 3.9). Model parameters are then estimated by optimizing the log-likelihood of the predicted covariance structure compared to the observed covariance. One major advantage of SEM is that it can be used to fit the same model while omitting some of the variance components. This can be used to assess the improvement of the fit by incorporating a genetic component in the model. Models can be compared based on their log-likelihood, or information criteria that additionally incorporate model complexity, such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). SEM can also be used to estimate heritability for categorical trait using the liability threshold model. It assumes that each category reflects an imprecise measurement of the underlying normal distribution (Rijdsdijk et al., 2002).

The increased availability of large genotyped cohorts facilitated the development of methods to estimate heritability from genotyping data directly (Speed et al., 2012; Yang et al., 2011). However, in contrast to family-based methods they only estimate the effect of genetic variation that is measured (or at least tagged by) the genotyping data. This usually excludes rare variants and structural variation. Also, these methods require much larger sample sizes than twin modelling (at least several thousand).

In this thesis, I estimated heritabilities using SEM implemented in the R package *metas*. I fitted four different models modelling different variance components: the full *ACE* model and *AE*, *CE* and *E* models. I selected the best fitting model based on the AIC. If not stated otherwise, all heritability estimates were adjusted for age and gender.

Figure 3.9 Twin modelling. Twin modelling exploits the fact that monozygotic (MZ) twins share (almost) 100 % and dizygotic (DZ) twins share 50 % of their genetic variation (A), but in both cases siblings are exposed to the same common environmental factors (C), in utero and when growing up. Additionally, there are unique environmental factors that are different for every individual (E). Comparing MZ and DZ twins allows to differentiate these three variance components for various phenotypes.



3.4.2 Univariate analysis

In most epidemiological studies researchers aim to identify biomarkers for an outcome of interest with many individual variables, such as SNPs or metabolites. These hypothesis-free approaches are referred to as genome-wide association studies (GWASs) and metabolome-wide association studies (MWASs), respectively.

In this thesis, I used regression models to test for bivariate dependencies. Regression models facilitate adjustment for confounding factors and can be applied to a wide range of variables, including Gaussian distributed, dichotomous, and categorical. In general, regression models aim to approximate one variable, the outcome or dependent variable y , by a linear combination of independent variables x_1, \dots, x_n . To this end, the independent variables are weighted by regression coefficients β_1, \dots, β_n and the weightings are fitted to maximise the likelihood of the model by minimising the random error term ϵ .

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n + \epsilon \quad (3.3)$$

This concept is further extended by random effect models, which assume that independent variables affect the outcome depending on a group structure of samples, such as different populations, study centres or families. These 'structuring' variables are thought of as random observations drawn from a larger entirety, such as a set of families drawn from all families or a set of study centres drawn from all (theoretical) study centres. These effects are, thus, called random effects and the according models random effect models. Regression models incorporating both random and non-random, i.e. fixed, effects are called mixed effects models.

For this thesis, association studies, including MWASs, were conducted by calculating regression models as implemented in the R-packages `stats` and `lme4` (R Core Team, 2016; Bates et al., 2015). Models were usually adjusted for age, gender and BMI. When analysing data from the TwinsUK cohort a random intercept was added per family to account for the twin structure.

3.4.2.1 Multiple testing

Significance of associations is assessed using p -values, which express the probability that the observed association occurs due to random chance. Usually findings with a p -value of 5% are further considered, i.e. a 5% chance of false positive discoveries is accepted. However, when conducting association studies using high-dimensional data, not one but many – often thousands of – hypotheses are tested simultaneously. Accepting an error rate of 5% within this group (or family) of thousands of statistical tests leads to an inflation of false discoveries. Thus, association studies presented in this thesis were corrected for multiple testing using the Bonferroni method. The p -values were multiplied by the number of tests that were performed within this family of tests, thus controlling the family-wise error rate (FWER). The FWER is the probability of observing at least one false positive discovery within this family of tests.

3.4.2.2 Meta-analyses

Association studies are often conducted in several different cohorts to achieve larger sample sizes and validate results independently of potential study effects. Meta-analyses are used to combine effect estimators and p -values across studies. Thereby results from individual studies are weighted by the accuracy of their effect estimator. Various weighting methods are available, for instance the inverse variance method, which weights each study by the inverse of the variance of the respective effect estimator (Borenstein et al., 2010).

In general, there are two different concepts to meta-analysis: Fixed-effects meta-analysis assumes that the true effect estimator is the same for all studies and aims to identify this single true effect. In contrast, random-effects models assume variation of the effect estimators between studies and aims to estimate their average (Borenstein et al., 2010). Due to heterogeneity of individual study populations, with respect to age, geographic location and genetic make-up, random-effects meta-analysis might be closer to the truth in many cases, however the random effects are harder to estimate and results were shown to be inaccurate, particular in settings with a limited number of studies (Guolo et al., 2015).

As a maximum of three cohorts were meta-analysed and all populations used in this thesis are European with similar demographics (Table 3.1), I used inverse variance fixed-effects meta-analysis throughout this thesis, as implemented in the R package *meta* (version 4.3).

3.4.2.3 Genome-wide association studies

GWASs are used to analyse genetic influences on traits of interest, such as metabolite levels, in a hypothesis-free manner. Prior to analysis, SNPs were tested for Hardy-Weinberg equilibrium using unrelated individuals, which were randomly drawn from the study population, with the *Plink* software suite (version v1.90b3.38). SNPs with $P < 10^{-6}$ were considered to deviate significantly from Hardy-Weinberg equilibrium and were, thus, excluded from further analysis. GWASs were conducted by fitting linear mixed models using *GEMMA* (version 0.94.1) (Zhou et al., 2012), adjusting for age and gender as fixed effects. Additionally, sample kinship was calculated from genotyping data using the *GEMMA* software suite and sample relatedness was

added to the model as a random effect. The score test implemented in GEMMA was used to calculate p -values for all associations. A multiple testing-corrected significance threshold was calculated by dividing the genome-wise significance threshold of 5×10^{-8} by the number of analysed phenotypes.

3.4.3 Pathway enrichment

Pathway enrichment considers the structure of biological data by analysing groups of related metabolites or genes. Therefore, it can be used to identify patterns in the results of an association study. Moreover, it can be used to identify affected pathways, even though its components might not be associated with the trait individually (see section 1.3.2.1).

There are several methods available to test for over-representation of certain pathways. As a start one can divide the variables in two groups '*significant*' and '*non-significant*' depending on their association with the phenotype and then test for enrichment of certain groups (pathways) amongst the significant variables by comparing the observed numbers with expected values drawn from a hyper-geometric distribution. Even though this is a valid method for pathway enrichment, it requires separation of variables in two groups and therefore ignores the ordering amongst them and associations below the (arbitrary) significance threshold.

To overcome this problem Subramanian et al. (2005) introduced the gene set enrichment analysis (GSEA) algorithm, a non-parametric approach which uses the ranking of variables according to their association with the phenotype of interest to calculate an enrichment score. The significance of the enrichment score is then assessed by estimating empirical p -values from random permutations of the data. This approach was the first to overcome the limitation of an arbitrarily chosen threshold. Kim et al. (2005) further improved the method by developing the parametric analysis of gene set enrichment (PAGE) algorithm using a parametric statistic, which fully takes into account the strength of the correlation of each variable with the outcome.

All pathway enrichments in this thesis were performed using the PAGE algorithm implemented in the R-package `piano` (Väremo et al., 2013). Enrichment tests are performed for each given pathway while taking into account the sign of the associations. Statistical significance of all enrichment analyses were assessed using empirical p -values estimated from a background distribution of 10,000 random permutations of the variable labels, and multiple testing correction was used to correct for the number of tested pathways.

3.4.4 Multi-variable regression (LASSO)

Another way to consider interdependencies of multiple predictors, e.g. metabolites, are multi-variable regression models. These are useful to assess the correlations of a phenotype with multiple different metabolites independently of each other. Using a panel of predictors can help to increase the predictive performance compared to single biomarkers, if the different

variables contain complementary information. However, such models are prone to over-fitting, particularly in high-dimensional settings with more variables than observations. Also, they tend to consist of many predictors, most of which have very small influence on the prediction.

To address both problems Tibshirani (1994) introduced a modified regression method, the least absolute shrinkage and selection operator (LASSO). The LASSO penalises the sum of regression coefficients, and therefore only adds variables to the model that increase the predictive performance. It thus, implicitly performs a variable selection by setting regression coefficients of variables that do not considerably increase predictive performance to $\beta_j = 0$, i.e. removing them from the model. Mathematically speaking, instead of minimising the predictive error of the model

$$\sum_{i=1}^N \left(y_i - \sum_j \beta_j x_{ij} \right)^2 \quad (3.4)$$

with the outcome variable y and the independent variables x_j , weighted by the regression coefficients β_j , it adds a penalty term and minimises

$$\sum_{i=1}^N \left(y_i - \sum_j \beta_j x_{ij} \right)^2 + \lambda \sum_j |\beta_j| \quad (3.5)$$

The tuning parameter λ determines the strength of the regularisation. λ can vary from $\lambda = 0$, which resembles a normal regression, to $\lambda = \infty$, which will lead to an empty prediction model. In practice a series of λ -values is tested and their predictive performance is evaluated using a cross-validation (see below) and the λ that minimises the predictive error or an information criterion is chosen for the final model. Here, I used the R-package `glmnet` to fit LASSO models. The function `cv.glmnet` was used to estimate the optimal regularisation parameter λ using cross-validation.

The main advantage of LASSO over other machine learning approaches is that variables that contribute to the model can be immediately identified by their non-zero regression coefficient. In contrast, most other machine learning methods, such as support vector machines (SVMs), differentiate classes of observations (such as cases and controls) by defining hyper-variables, that consist of linear (or even non-linear using kernels) combinations of the original variables. Other approaches, such as random forests, do not even use a single classifier but an ensemble of multiple classifiers, each of which uses different sets of variables. Finally, neuronal networks are almost complete black boxes. While these machine learning approaches are immensely useful if the purpose is only the optimal classification, biological interpretation of the results is challenging in case of SVMs and random forests and virtually impossible for neuronal networks. Similarly, most dimensionality reduction methods usually define meta-variables based on (linear) combinations of the original variables, which reduces correlation between variables as well as the burden of multiple testing. However, interpretation of these abstract combinations of many biological entities is hardly possible. As I was aiming to gain knowledge about biological processes underlying ageing, I did not use such methods to reduce dimensionality or train classifiers.

3.4.5 Imputation

A disadvantage of multivariate methods, such as the LASSO but also other multivariate methods described below, is that they rely on full data matrices. As most real-world datasets are not complete, amongst others due to incomplete questionnaires, limits of detections or failed measurements, imputation of missing values is required to use these methods, which introduces further complexity.

Here, I imputed missing values using the R package `mice` (Buuren et al., 2011). I used the `norm` method, which calculates Bayesian linear regression models, predicting each variable based on all other variables. Missing values are then imputed as the sum of the prediction based on the regression models and a random error term depending on the residual error of the models. For categorical variables, logistic regression models are used. `Mice` imputation, thus, preserves the multivariate covariance structure of variables.

3.4.6 Assessing predictive performance

If the aim of a study is to predict a phenotype or disease of interest, the predictive performance must be assessed to rate the usability of the predictive model. Depending on the required outcome different measures are commonly used.

3.4.6.1 Binary outcomes

In case of binary outcomes one can count the number of true positives (TPs) – the number of correctly predicted cases, true negatives (TNs) – the number of correctly predicted controls, false positives (FPs) – the number of controls that were incorrectly predicted as cases, and false negatives (FNs) – the number of cases that were incorrectly predicted as controls (Figure 3.10).

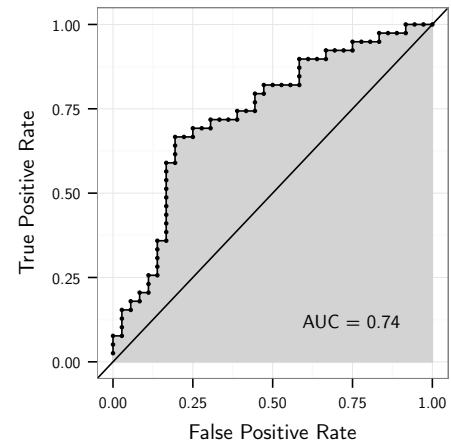
From these one can calculate the sensitivity (or true positive rate (TPR)) as $TP/(TP + FN)$, which measures the performance of the method in detecting cases, and the specificity (or true negative rate (TNR)) as $TN/(TN + FP)$, which measures its performance in detecting controls. In many cases, algorithms do not predict the binary outcome, but a continuous measure that represents the confidence of the classification as case. Consequently, one has to choose one cut-off to classify samples in predicted cases and controls. For each of these cut-offs one can count the number of true and false predictions and calculate TPR and false positive rate (FPR) from these. The receiver operating characteristic (ROC) curve summarises this spectrum of predictions by plotting TPR and FPR of each possible threshold against each other. The area under the curve (AUC) gives an overall measure of predictive performance, where a value of 1 indicates a perfect prediction and a value of 0.5 corresponds to random guessing (Figure 3.10).

Figure 3.10 Assessment of statistical predictions. (a) When assessing the performance of a predictor to predict a binary outcome, one can initially count the number of true – true positive (TP) and true negative (TN) – and false – false positive (FP) and false negative (FN) predictions. Ratios of these counts over the total number of positive and negative observations are called true positive rate (TPR) and true negative rate (TNR), respectively. Ratios of the counts over the total number of positive and negative predictions are called positive predictive value (PPV) and negative predictive value (NPV), respectively. (b) When predicting a binary outcome with a continuous predictor, one can calculate these measures for all possible cut-off points and summarise the results in a receiver operating characteristic (ROC) curve, where each point corresponds to the true positive rate (TPR) and true negative rate (TNR) of one cut-off. The area under this curve (AUC) can be calculated by integration and provides an overall measure of prediction performance.

(a) Confusion table

		Predicted		
		Pos. (P_p)	Neg. (N_p)	
Observed	Pos. (P_o)	TP	FP	TPR = $\frac{TP}{P_o}$
	Neg. (N_o)	FN	TN	TNR = $\frac{TN}{N_o}$
		PPV = $\frac{TP}{P_p}$ NPV = $\frac{TN}{N_p}$		

(b) ROC curve



3.4.6.2 Continuous outcomes

If the predicted variable is continuous one can calculate the total prediction error – the residual sum of squares (RSS) – as

$$\text{RSS} = \sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2,$$

for n samples, where y_i and \hat{y}_i are the observed and predicted values of the variable y for sample i , respectively. Averaging the total error gives the mean squared error (MSE), which is commonly used for fitting LASSO models. However, the MSE is on the scale of the predicted variable and can, thus, not easily be compared across studies. The coefficient of determination R^2 normalises the RSS

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

to express the predictive performance as proportion of variance explained by the predictor. The total sum of squares (TSS) is defined as follows, equal to the sample variance

$$\text{TSS} = \sum_i (y_i - \bar{y})^2$$

3.4.6.3 Cross-Validation

A common problem of statistical prediction is over-fitting of data. Over-fitting occurs when the prediction algorithm learns structures in the data that cannot be generalised. Consequently, the predictive performance will be over-estimated in the original dataset, but poor for other datasets. For an unbiased assessment of its performance, a predictor has to be tested in an independent dataset, that was not used for training. If no external validation sets are available, one can split the dataset in two to train the algorithm on the first and test its performance on the second split. A cross-validation uses the same principle and repeats it iteratively until each sample was once used for the testing. Usually 10 iterations, or folds, each of them using 90 % of the data for training and the remaining 10 % for testing, are performed. A special case of cross-validation is the leave-one-out (LOO) validation, where in each fold a single sample is left out for testing and all other samples are used for training.

Using the left-out dataset for testing, one can compute the same statistics as described above. Summary statistics can then be derived as mean from all folds. When the RSS is calculated in a cross-validation scenario it is called predicted residual sum of squares (PRESS). The corresponding R^2 value is referred to as P^2 .

3.4.7 Graphical models

To model multivariate dependencies in high-dimensional datasets, I used GMs (Lauritzen et al., 1989). These model conditional dependencies between variables, thus overcoming the problem of spurious correlations (see section 1.3.3.2 for details).

One of the most famous approaches for inference of sparse graphical models is the graphical LASSO (Friedman et al., 2008; Mazumder et al., 2012). Like the LASSO, it penalises the sum of partial correlations, thus enforcing sparse graphical models and avoiding over-fitting of the data. Just as for the LASSO selecting a regularisation parameter (in this context often referred to as ρ) is challenging. It can be chosen based on information criteria, such as the AIC or BIC, however these tend to produce very dense graphs. Alternatively, one can use different cross-validation or bootstrapping procedures, which produce sparser graphs, but are computationally expensive for large datasets (Liu et al., 2010).

Instead of enforcing sparseness by regularisation, the GeneNet algorithm (Schäfer et al., 2005) utilises a different regularisation approach to estimate the covariance matrix as good as possible in high-dimensional settings. This approach allows us to determine the amount of regularisation required analytically and does therefore not rely on elaborate optimisation procedures (Ledoit et al., 2003). To avoid dense networks due to very small non-zero entries in the precision matrix, a false discovery rate (FDR) is estimated for every edge, i.e. every partial correlation, by estimating a background distribution from the entirety of all partial correlations. Edges of the model can then be selected by applying a significance threshold.

For this thesis, I calculated GGMs using the GeneNet approach due to its computational efficacy, straight-forward choice of the regularisation parameter, and its proven applicability to metabolomics data (Krumstiek et al., 2011; Valcárcel et al., 2011). I used the `estimate.lambda` function of the R-package `corpcor` to calculate the regularisation parameter λ , the function

`pcor.shrink` to estimate shrinkage partial correlations, and `fdrtool` to estimate a FDR for each partial correlation (Schäfer et al., 2005). Shrinkage partial correlations with $\text{FDR} < 5\%$ were considered significant and used as edges of the model.

3.4.8 Mixed Graphical Models

While GMs are well-defined and often used in the context of multivariate Gaussian-distributed data, their application on less well-distributed data lacks behind. Some recent developments facilitate the inference of GMs from not Gaussian-distributed datasets and more interestingly mixed datasets, comprising both continuous and categorical variables. In chapter 9, I present the first application of a mixed graphical model approach to a large dataset of omics-markers combined with clinical phenotypes.

To this end, I used the graphical random forest (GRaFo) method (Fellinghauer et al., 2013) with the complementary pairs stability selection (CPSS) modification (Shah et al., 2013) to infer mixed graphical models (MGMs). For each variable, all the remaining variables are ranked according to their conditional dependence, which I quantified as the random forest variable importance. Consequently, two ranks are calculated for each pair of variables x_1 and x_2 : one based on the variable importance of x_2 for the prediction of x_1 , and the other based on the importance of x_1 for the prediction of x_2 . The maximum (i.e. worse) of these two ranks is then used as rank of the pair.

This procedure is repeated for 100 random subsets of the data, each of them containing the half of all samples, and their respective complementary sets containing the other half of the samples. This results in 100 pairs of lists, each of them ranking all possible edges (i.e. pairs of variables) according to their conditional dependencies. Each of these lists can be considered permutations of the underlying graphical model, where the highest-ranking pairs of variables are connected by edges.

To combine these permutations in one stable MGM CPSS (Shah et al., 2013) is used, which also controls the FWER of edges, thus facilitating an informed decision on a cut-off for edge inclusion. To this end, an edge is included in one complementary pair if it was amongst the top-scoring edges in both complementary sets and edges included in at least 80 % of all complementary pairs are included in the final model. One can calculate the number of edges to include in each of the subsamples to ensure $\text{FWER} < 5\%$ in the final network by using the formula 3.6 proposed by Fellinghauer et al. (2013)

$$\mathbb{E}|\hat{S}_{n,\tau}^{CPSS} \cap N| \leq \frac{1}{2\tau - 1} \frac{q^2}{p}, \quad (3.6)$$

where $\mathbb{E}|\hat{S}_{n,\tau}^{CPSS} \cap N|$ is the expected number of false positive edges, τ is the proportion of subsamples in which an edge has to be top-ranking to be included in the final model, q is the number of edges, and p the number of selected edges per subsample.

I implemented this method in R using the `randomForest` package for the assessment of conditional dependencies and the `plyr` and `Rmpi` packages for parallelisation to facilitate the inference of large models.

To summarise, conditional dependencies between variables are estimated using the random forest variable importance. Permutation samples are used to obtain a stable ranking of edges and CPSS is used to include edges that pass a FWER of 5% in the final graphical model.

3.4.9 Graph analysis

The whole field of graph theory, a specialisation of maths, is devoted to the analysis of graphs. Hence, various measures were proposed to describe the topology of networks and topological features of nodes, trying to identify highly connected, central nodes. These are thought to be key players in the system, connecting several modules and controlling network fluxes. They were shown to be of particular importance for many diseases and survival of the organism (Barabási et al., 2004; Joy et al., 2005; Yu et al., 2007).

3.4.9.1 Topological properties of nodes

The most straight-forward measure of node centrality is the degree, which is the number of edges that connect a node with the rest of the model. In a directed graph one can distinguish the in-degree, the number of incoming edges, and the out-degree, the number of outgoing edges. The degree provides a basic measure of node connectivity. The (local) clustering coefficient measures how densely the neighbourhood of a node is connected. It is defined as the proportion of edges between the neighbours of a node that are present in the network. The global clustering coefficient summarises this property for the overall network and can be calculated as the number of connected triangles over the number of paths of length 2. In contrast to the clustering coefficient, the betweenness centrality describes the importance of a node for the connection of all other pairs of nodes. It is defined as the number of shortest paths between all pairs of nodes, that contain the node of interest (Barabási et al., 2004) (Figure 3.11). All these measures can be generalised to weighted networks in which edges are not only present or absent but present edges are weighted by their importance (Langfelder et al., 2008).

3.4.9.2 Modules

Real-world networks often consist of densely connected modules, or clusters, that represent functional units within the network (Fortunato, 2010). Several methods have been developed to identify modules of nodes, which are jointly affected by the condition of interest. Two publicly available examples are the Cytoscape plugin jActiveModules (Ideker et al., 2002) and the R package BioNet (Beisser et al., 2010). In this thesis, all graphs were analysed using the R-package igraph (Csardi et al., 2006). Module (or cluster) identification was conducted using the function `cluster_optimal`, which identifies the partition of the graph into modules that maximises the graph modularity. Thereby the modularity Q is defined as

$$Q = \frac{1}{2m} \cdot \sum_{i,j} \left(A_{ij} - \frac{k_i \cdot k_j}{2m} \cdot I(c_i, c_j) \right) \quad (3.7)$$

where m is the number of edges in the graph and i and j nodes with degrees k_i and k_j , and cluster assignments c_i and c_j , respectively (Clauset et al., 2004). $I(c_i, c_j)$ is an indicator function that returns 1 if i and j are in the same cluster and 0 otherwise. A is the adjacency matrix and A_{ij} the edge-weight of the edge connecting i and j , where $A_{ij} = 0$ indicates absence of the edge. Thus, maximising Q will identify clusters with the maximal number of within-cluster edges compared to the number of expected within-cluster edges if modules were defined by random (Clauset et al., 2004).

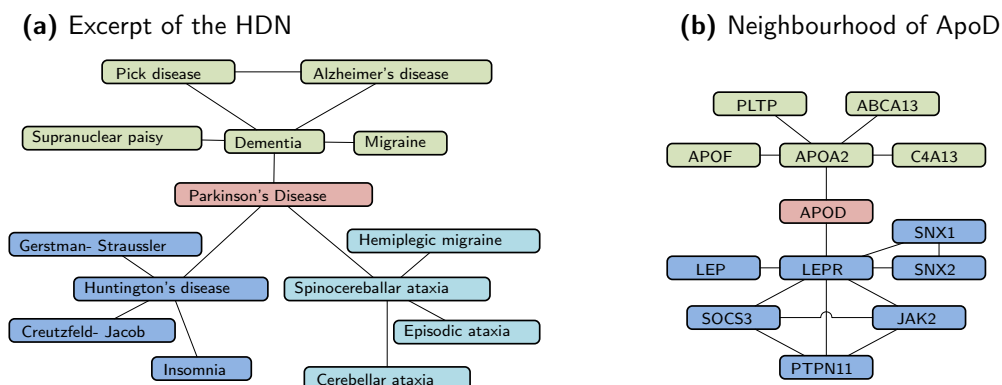
3.4.9.3 Small-world networks

In most biological networks, average path lengths between nodes are usually short despite their high modularity. Such networks are referred to as small word networks. The small word index of a network quantifies this property by comparing its clustering coefficient and average shortest path lengths with a Erdős-Rényi random graph (Humphries et al., 2008). It is defined as

$$S_g = \frac{C_g / C_{rand}}{L_g / L_{rand}}, \quad (3.8)$$

where C_g and C_{rand} are the global clustering coefficients of the graph g and a random graph, respectively. L_g and L_{rand} are their respective average shortest paths lengths between all pairs of nodes. A network with $S_g > 1$ is considered a small-world network.

Figure 3.11 Topological properties of nodes in graphs. (a) Exemplary, an excerpt of the human disease network (HDN) (Goh et al., 2007) is shown. Nodes represent diseases that are connected if they are associated with the same gene. Parkinson's disease connects three isolated disease clusters (as indicated by colours), thus having a low clustering coefficient (0 %) and high betweenness centrality (72 %). (b) The close neighbourhood of the ApoD protein in a PPI network from STRING DB (Franceschini et al., 2013) using only experimentally confirmed interactions, is used as a second example. ApoD connects two clusters and is, despite the low degree (2) and clustering coefficient (0 %), a central node, as indicated by the betweenness centrality of 53 %. In contrast, LEPR is central within the blue cluster and has a high degree of 7, and clustering coefficient of 14 %. (Picture adopted from Zierer et al. (2015))



CHAPTER 4

Metabolic markers of leukocyte telomere length

In this chapter, I used leukocyte telomere length as marker of biological ageing and associated it with circulating metabolite levels. By doing so I identified five metabolites associated with telomere length and additionally with ageing-related phenotypes.

This chapter has been published in *Ageing* (Zierer et al., 2016a). The published version is attached in appendix G.2.

Telomeres are repetitive DNA sequences that are located at the end of the chromosomes to protect them from degradation and end-to-end-fusion (see section 1.1.2.1). In somatic cells telomeres shorten with every cell cycle due to the incomplete replication of the 3' end of the template strand and were thus proposed as a potential cause of cell senescence (Harley et al., 1992) and biomarkers of ageing (Mather et al., 2011). Consequently, leukocyte telomere length (LTL) has been associated with chronological age as well as a multitude of age-related diseases (section 1.1.2.1), however it is still unclear how telomere shortening affects health.

In this chapter, I aimed to investigate metabolic changes that co-occur with the shortening of telomeres to identify potential pathways of biological ageing. I then further analysed the identified metabolites with respect to age-related phenotypes as well as changes of gene expression and DNA methylation.

4.1 Methods

Study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins recruited as volunteers without selecting for any particular disease or trait (see section 3.1.1). In this study, I analysed data of 3511 female twins who had complete data for LTL and metabolomics profiling.

I replicated significant associations in 904 females of the KORA F4 cohort (see section 3.1.2) with measures of LTL and blood metabolic profiles available.

4.1.1 Leukocyte telomere length

LTL was measured as a relative measure of telomere repeat length over a single copy gene as described in section 3.2.4. Relative LTL was on average 3.72 in TwinsUK and 1.85 in KORA (Figure 4.1). Measurements were inverse normalised in both cohorts.

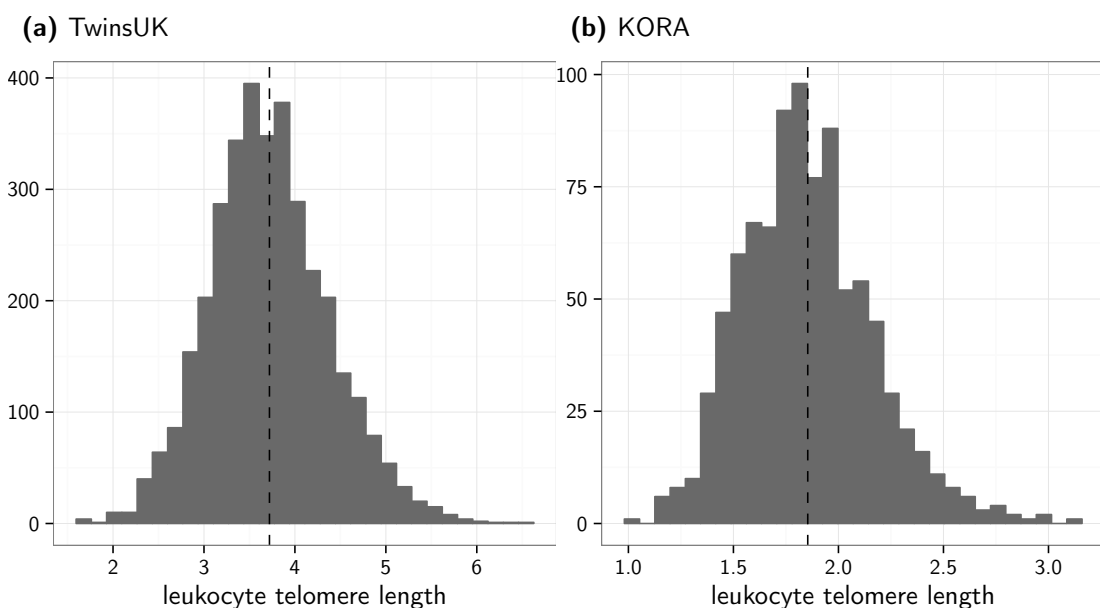
4.1.2 Metabolomics measurements

Metabolomics data for both cohorts was measured using the V3 platform of Metabolon Inc., as described in section 3.3.1.1. In this study, I analysed 280 metabolites with known chemical identity (Table A.1).

4.1.3 Ageing phenotypes

Several age-related phenotypes were additionally analysed to further explore the association of the observed metabolic changes with health. Renal function was quantified as eGFR (see section 3.2.3 for details). Liver function was assessed by measuring serum GGT and ALT concentrations (section 3.2.5); both measures were inverse normalized prior to analysis. Lung function was measured as FEV1 as described in section 3.2.6. Furthermore, diastolic and systolic blood pressure (DBP and SBP), BMI (section 3.2.1) and serum high density lipoprotein (HDL) cholesterol levels were measured during clinical visits of the study participants.

Figure 4.1 Telomere length in TwinsUK and KORA. Relative measures of leukocyte telomere length (LTL) were analysed from (a) 3511 individuals from TwinsUK and (b) 904 individuals from KORA. Histograms show the distribution of LTL in the respective populations, dashed lines indicate the mean values.



4.1.4 Gene expression

RNA abundance was measured in lymphoblastoid cell lines (LCLs) of 778 female individuals from the TwinsUK cohort using the Illumina Human HT-12 V3 BeadChip (Illumina, San Diego, CA, USA) as part of the Multiple Tissue Human Expression Resource (MuTHER) project as previously described (Grundberg et al., 2012). I selected 30 probes from the GGT and PLA2 genes. Probes were adjusted for batch effects by linear models and residuals were inverse normalized prior to analysis.

4.1.5 DNA methylation

DNA was extracted from whole blood, bisulfite converted and subsequently analysed using the Infinium 450K kit (Illumina, San Diego, CA, USA) as previously described (Tsai et al., 2015). The beta mixture quantile dilation (BMIQ) method was performed to correct for technical variation (Teschendorff et al., 2013). Measurements were inverse normalized and then adjusted for batch effects, family structure and cell counts (PlasmaBlast, CD8+CD28-CD45RA-T cells, naive CD8 T cells, CD4+ T cells, Natural Killer cells, monocytes, and granulocytes) using linear models.

4.1.6 Statistical analysis

Correlations between metabolite levels and LTL were calculated using linear mixed models, correcting for age, BMI, and family relatedness as random intercept (as described in section 3.4.2). Results were corrected for 280 tested metabolites using the Bonferroni correction ($p < 1.8 \times 10^{-4}$) (section 3.4.2.1).

I replicated five Bonferroni-significant metabolites in the KORA F4 cohort (section 3.1.2). The data was consistently normalized and analysed in both cohorts and results were meta-analysed using inverse variance fixed effect meta-analysis (see section 3.4.2.2). I estimated the power of the replication cohort using the R package *pwr* (version 1.1), which implements power estimation according to Cohen (1988).

To identify non-redundant associations of the metabolites with LTL, I fitted a multivariate LASSO model (Tibshirani, 1994) incorporating all metabolites passing the Bonferroni cut-off together with age and BMI (see section 3.4.4). The predictive performance of the model was then compared to a baseline-model, containing age and BMI only, by calculating the PRESS and P^2 statistics using a LOO cross-validation (details in section 3.4.6).

Subsequently, I aimed to further explore the relationship of the identified metabolic profile with biological ageing. To this end, I used linear mixed models to test for association of the previously identified metabolites with ageing-related phenotypes. All models were adjusted for age, BMI and family relatedness as described before. The lung function parameter FEV1 was additionally adjusted for height, as suggested in the literature. I replicated the associations with lung function parameters in KORA, adjusting for the same covariates.

4.2 Results

I analysed the associations of 280 fasting blood metabolites with LTL in 3511 women from the TwinsUK cohort (Table A.1) and replicated the findings in the KORA cohort. The demographic characteristics of the study populations are presented in Table 4.1.

Five metabolites were significantly negatively associated with LTL after adjustment for potential confounding factors and correction for multiple testing (Table 4.2): the two lipids 1-stearoylglycerophosphoinositol ($\beta[95\%CI] = -0.07 [-0.10 : -0.04]$ change in metabolite z-score per change in LTL z-score, $p = 1.6 \times 10^{-5}$) and 1-palmitoylglycerophosphoinositol ($\beta[95\%CI] = -0.08 [-0.12 : -0.04]$, $p = 1.6 \times 10^{-5}$), the two gamma-glutamylamino acids gamma-glutamyltyrosine ($\beta[95\%CI] = -0.08 [-0.11 : -0.05]$, $p = 2.5 \times 10^{-6}$) and gamma-glutamylphenylalanine ($\beta[95\%CI] = -0.07 [-0.10 : -0.04]$, $p = 1.7 \times 10^{-5}$), and the xenobiotic 4-vinylphenol sulfate ($\beta[95\%CI] = -0.07 [-0.10 : -0.03]$, $p = 1.4 \times 10^{-4}$). All five metabolites showed the same effects with similar effect sizes in 904 female individuals from the KORA F4 cohort, though they did not reach significance level. All metabolites remained Bonferroni-significant ($p < 1.8 \times 10^{-4}$) after meta-analysis (Figure A.1).

Table 4.1 Population characteristics. To identify metabolic markers of leukocyte telomere length, I analysed 3511 female twins as discovery and 904 females from the KORA cohort as replication cohorts. Population characteristics are shown here.

	TwinsUK	KORA
<i>N</i>	3511	904
<i>MZ:DZ:Singletons</i>	1654:1360:497	0:0:904
<i>Age, years</i>	53.6 \pm 13.6	60.5 \pm 8.8
<i>LTL</i>	3.72 \pm 0.67	1.85 \pm 0.31
<i>BMI, kg/m²</i>	26.21 \pm 5.14	27.87 \pm 5.25
<i>FEV1, L</i>	2.60 \pm 0.61	2.79 \pm 0.50
<i>HDL, mmol/L</i>	1.71 \pm 0.48	
<i>DBP, mmHg</i>	78.01 \pm 10.68	
<i>SBP, mmHg</i>	126.71 \pm 18.20	
<i>ALT, IU/L</i>	27.63 \pm 17.07	
<i>GGT, U/L</i>	28.36 \pm 25.44	
<i>eGFR, mL/min/1.73m²</i>	83.78 \pm 17.07	
<i>smoking(non:former:current)</i>	1905:1134:447	

Table 4.2 Metabolites significantly associated with leukocyte telomere length. Five metabolites were significantly associated with leukocyte telomere length in the discovery cohort (TwinsUK) and the meta-analysis.

Metabolite	Pathway	TwinsUK			Kora			Meta		
		β	[95 %CI]	p-value	β	[95 %CI]	p-value	β	[95 %CI]	p-value
<i>gamma-glutamyltyrosine</i>	Peptide	-0.09	[-0.12 : -0.05]	3.4×10^{-6}	-0.05	[-0.12 : 0.02]	0.02	-0.08	[-0.11 : -0.05]	2.5×10^{-6}
<i>1-stearoylglycerophosphoinositol</i>	Lipid	-0.09	[-0.13 : -0.05]	1.4×10^{-6}	-0.00	[-0.07 : 0.07]	0.07	-0.07	[-0.10 : -0.04]	1.6×10^{-5}
<i>1-palmitoylglycerophosphoinositol</i>	Lipid	-0.08	[-0.13 : -0.04]	7.4×10^{-5}	-0.07	[-0.14 : 0.01]	0.01	-0.08	[-0.12 : -0.04]	1.6×10^{-5}
<i>gamma-glutamylphenylalanine</i>	Peptide	-0.08	[-0.12 : -0.04]	2.7×10^{-5}	-0.04	[-0.11 : 0.02]	0.02	-0.07	[-0.10 : -0.04]	1.7×10^{-5}
<i>4-vinylphenol sulfate</i>	Xenobiotics	-0.08	[-0.12 : -0.04]	7.4×10^{-5}	-0.03	[-0.10 : 0.05]	0.05	-0.07	[-0.10 : -0.03]	1.4×10^{-4}

4.2.1 Independent associations with LTL

I fitted three multivariate LASSO models predicting LTL to assess the independence of these associations: The first model using age and BMI only, the second model using the five identified metabolites only, and the third using combining both. The model based on metabolites alone could not achieve the performance of the model based on age and BMI, however, combining them significantly improved the prediction in the combined model (Figure 4.2). All five metabolites were selected in the optimal LASSO model ($\beta \neq 0$), suggesting non-redundant associations with LTL. The coefficient of determination, a measure of goodness of fit, of the final model was estimated at 14.5% in a LOO validation.

4.2.2 Ageing phenotypes

Moreover, I found the five metabolites strongly associated with several age-related phenotypes, independently of chronological age (Table 4.3): Both lysolipids correlated with increased SBP (1-stearoylglycerophosphoinositol: $\beta[95\%CI] = 1.09 [0.56:1.61]$, $p = 5.3 \times 10^{-5}$ and 1-palmitoylglycerophosphoinositol: $\beta[95\%CI] = 1.10 [0.52:1.67]$, $p = 1.7 \times 10^{-4}$). Additionally, 1-palmitoylglycerophosphoinositol was associated with the serum concentration of GGT, a marker of liver function ($\beta[95\%CI] = 0.08 [0.03:0.12]$, $p = 1.0 \times 10^{-3}$).

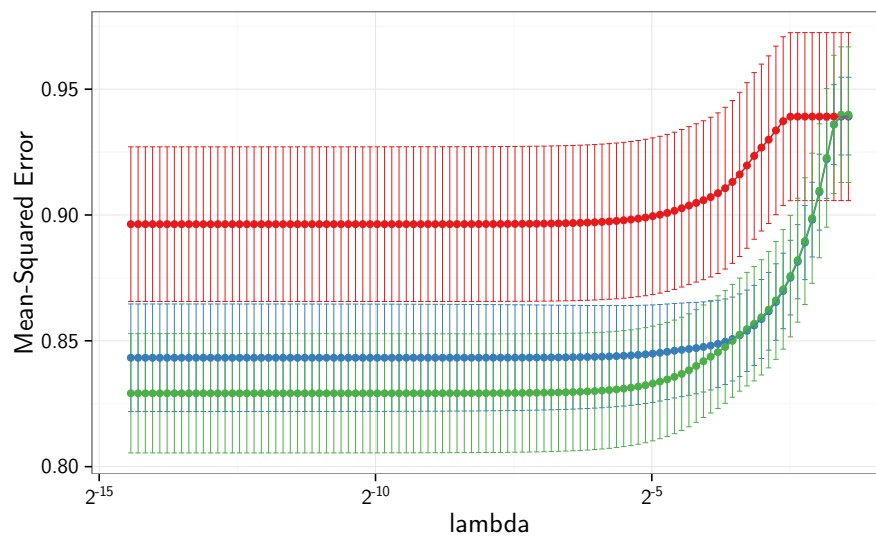
The two gamma-glutamylamino acids were strongly associated with eGFR, a marker of renal function (gamma-glutamyltyrosine: $\beta[95\%CI] = -1.65 [-2.19 : -1.11]$, $p = 1.6 \times 10^{-9}$ and gamma-glutamylphenylalanine: $\beta[95\%CI] = -2.24 [-2.74 : -1.73]$, $p = 3.1 \times 10^{-18}$), as well as both markers of liver function, GGT and ALT (GGT: $\beta[95\%CI] = 0.14 [0.10 : 0.19]$, $p = 5.4 \times 10^{-11}$ and $\beta[95\%CI] = 0.15 [0.10 : 0.19]$, $p = 3.2 \times 10^{-12}$, respectively; ALT: $\beta[95\%CI] = 0.11 [0.06:0.16]$, $p = 1.7 \times 10^{-5}$ and $\beta[95\%CI] = 0.10 [0.05:0.14]$, $p = 5.8 \times 10^{-5}$, respectively). Gamma-glutamylphenylalanine was additionally associated with lung function, measured as FEV1 ($\beta[95\%CI] = -0.03 [-0.05: -0.02]$, $p = 4.5 \times 10^{-6}$), and HDL cholesterol levels ($\beta[95\%CI] = -0.03 [-0.05: -0.02]$, $p = 1.1 \times 10^{-5}$). Moreover, the xenobiotic 4-vinylphenol sulfate was strongly associated with tobacco smoking ($\beta[95\%CI] = 0.24 [0.22:0.26]$, $p = 2.3 \times 10^{-102}$) and also with FEV1 ($\beta[95\%CI] = -0.02 [-0.04: -0.01]$, $p = 1.4 \times 10^{-3}$).

4.2.3 Expression of enzymes PLA2 and GGT

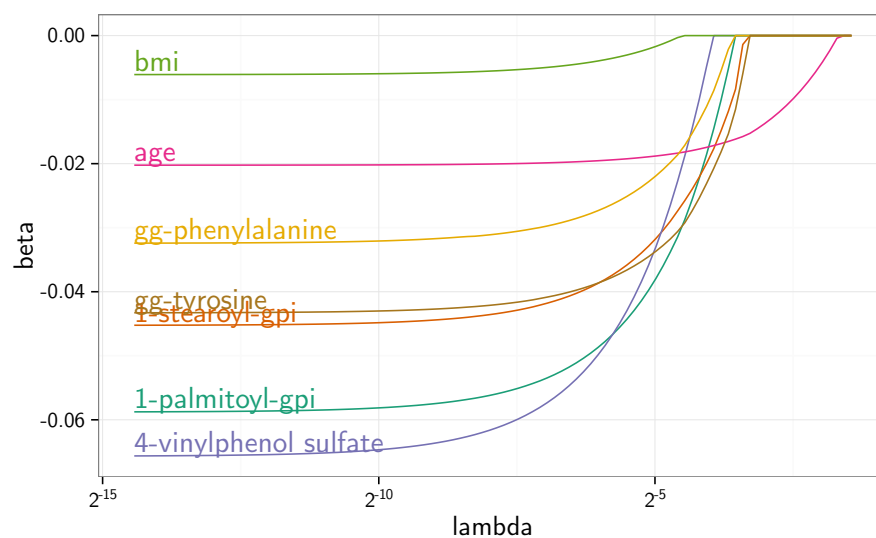
To further investigate mechanisms of biological ageing, I analysed the association of the five significant metabolites with gene expression levels of the related enzymes GGT and PLA2, in a subset of 753 individuals with transcriptomics data available. I found gamma-glutamyltyrosine positively associated with expression of *gamma-glutamyltransferase 1* (GGT1) and *gamma-glutamyltransferase-like 3* (GGTL3) (probes ILMN_2274240: $\beta[95\%CI] = 0.09 [0.02 : 0.15]$, $p = 1.0 \times 10^{-2}$ and ILMN_1786186: $\beta[95\%CI] = 0.07 [0.00 : 0.14]$, $p = 4.0 \times 10^{-2}$). Additionally, I found the expression of *phospholipase A2 group XV* (PLA2G15) positively associated with 1-stearoylglycerophosphoinositol (probe ILMN_1756910: $\beta[95\%CI] = 0.09 [0.01 : 0.16]$, $p = 2.0 \times 10^{-2}$ and probe ILMN_1798955: $\beta[95\%CI] = 0.08 [0.00 : 0.15]$, $p = 5.0 \times 10^{-2}$) as well as 1-palmitoylglycerophosphoinositol (probe ILMN_1756910: $\beta[95\%CI] = 0.08 [0.00:0.16]$, $p = 5.0 \times 10^{-2}$).

Figure 4.2 Prediction of leukocyte telomere length. (a) Three models were fitted to predict leukocyte telomere length (LTL): (1) using age and BMI as predictors (blue), (2) using five metabolites as predictors (red) and (3) using both combined as predictors (green). Their performance was measured as mean squared error (MSE) using cross-validation. (b) In the third model, all five metabolites contributed individually to the predictive performance.

(a) Performance of LASSO models to predict LTL



(b) Effects of all variables on prediction of LTL



4.2.4 Smoking-related DNA de-methylation

The metabolite 4-vinylphenol sulfate is known to be associated with several DNA methylation probes, possibly driven by tobacco smoking (Petersen et al., 2014; Zeilinger et al., 2013). I found one of these probes, cg19572487, being significantly associated with LTL ($\beta[95\%CI] = 0.10 [0.04 : 0.17]$, $p = 9.0 \times 10^{-3}$), smoking ($\beta[95\%CI] = -0.51 [-0.63 : -0.39]$, $p = 9.0 \times 10^{-16}$) and 4-vinylphenol sulfate levels ($\beta[95\%CI] = -0.05 [-0.09 : -0.02]$, $p = 1.0 \times 10^{-3}$) in our data. The probe is located on chromosome 17 in the *retinoic acid receptor alpha* (RARA) gene.

4.2.5 Integration of results

Combining those results, all five metabolites were consistently associated with accelerated biological ageing, i.e. shorter telomeres, higher blood pressure and poorer lung, liver and kidney functions. The metabolites were additionally correlated with the expression of related enzymes (Figure 4.3).

Figure 4.3 Metabolic signature of telomere length. Leukocyte telomere length (LTL) was significantly associated with five metabolites. Additionally, pairwise significant associations of these metabolites with several ageing-related phenotypes (round nodes) are shown: estimated glomerular filtration rate (eGFR), *alanine aminotransferase* (ALT), *gamma-glutamyltransferase* (GGT), high density lipoprotein (HDL) cholesterol, forced expiratory volume in one second (FEV1), systolic blood pressure (SBP) and smoking. Diamond-shaped nodes represent the expression levels of *gamma-glutamyltransferase* (GGT) and *phospholipase A2* (PLA2). Blue edges indicate significant negative associations, while red edges indicate positive associations.

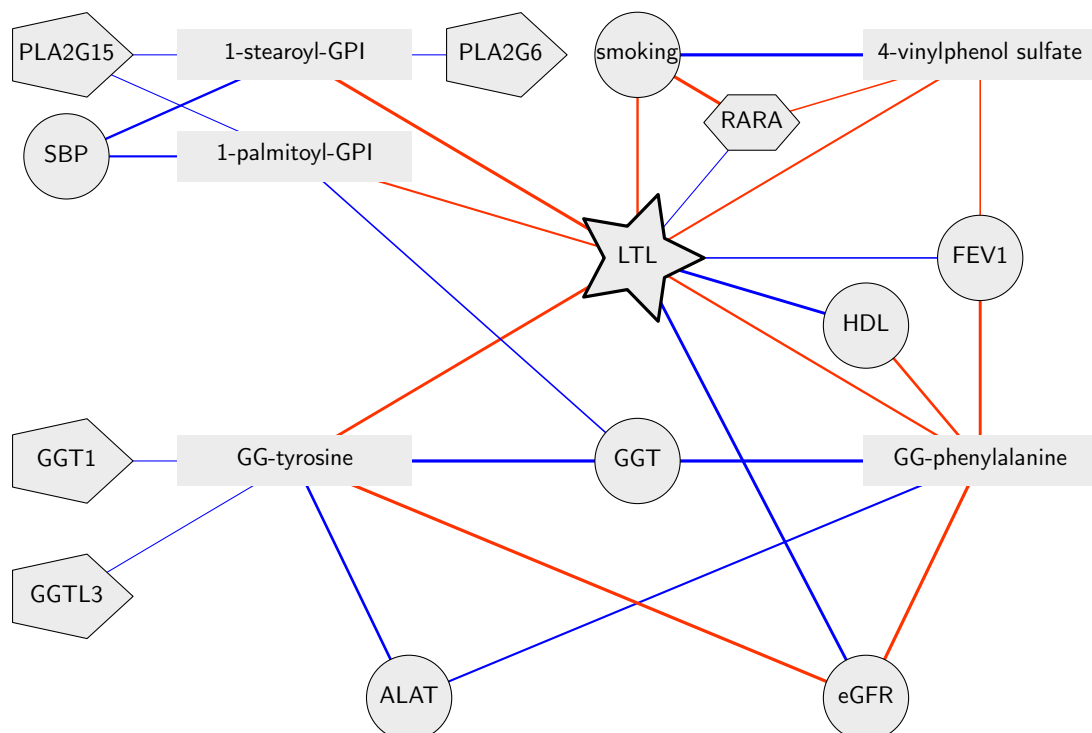


Table 4.3 Associations of telomere length and metabolites with age-related phenotypes.

The table shows significant associations of leukocyte telomere length and associated metabolites with various ageing phenotypes

	Phenotype	β [95 %CI]	p-value
<i>1-palmitoylglycerophospho-inositol</i>	SBP	1.10 [0.52: 1.67]	1.8×10^{-4}
	GGT	0.08 [0.03: 0.12]	1.0×10^{-3}
<i>1-stearoylglycerophospho-inositol</i>	SBP	1.09 [0.56: 1.61]	5.3×10^{-5}
<i>4-vinylphenol sulfate</i>	smoking	0.24 [0.22: 0.26]	1.4×10^{-101}
	FEV1	-0.02 [-0.04: -0.01]	1.4×10^{-4}
<i>gamma-glutamylphenyl-alanine</i>	eGFR	-2.24 [-2.74: -1.73]	3.1×10^{-18}
	GGT	0.15 [0.10: 0.19]	3.2×10^{-12}
	FEV1	-0.03 [-0.05: -0.02]	4.5×10^{-6}
	HDL	-0.03 [-0.05: -0.02]	1.1×10^{-5}
	ALT	0.10 [0.05: 0.14]	5.8×10^{-5}
<i>gamma-glutamyltyrosine</i>	GGT	0.14 [0.10: 0.19]	5.4×10^{-11}
	eGFR	-1.65 [-2.19: -1.11]	1.6×10^{-9}
	ALT	0.11 [0.06: 0.16]	1.7×10^{-5}
<i>LTL</i>	HDL	0.04 [0.02: 0.06]	2.5×10^{-6}
	eGFR	1.42 [0.82: 2.01]	2.8×10^{-6}
	smoking	-0.06 [-0.08: -0.03]	3.2×10^{-5}
	FEV1	0.03 [0.01: 0.05]	8.9×10^{-4}

4.3 Discussion

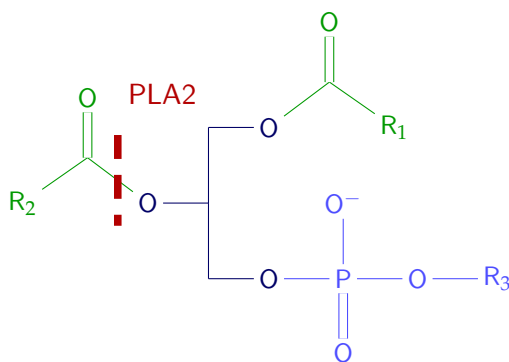
In the largest study of this kind, I searched for molecular markers and mechanisms involved in LTL regulation using a metabolomics approach. I identified five novel blood metabolites, namely gamma-glutamyltyrosine, gamma-glutamylphenylalanine, 1-stearoylglycerophosphoinositol, 1-palmitoylglycerophosphoinositol and 4-vinylphenol sulfate, independently associated with LTL with high statistical significance. These metabolites belong to three different classes: lysolipids, gamma-glutamylamino acids and xenobiotics, which I will discuss here.

4.3.1 Lysolipids

Lysolipids are produced from glycerophospholipids by the enzyme PLA₂, which releases one of the fatty acids from the glycerol backbone (Figure 4.4) (Dennis, 1994). Glycerophospholipids were previously found to be positively correlated with LTL (Zhao et al., 2014), while in this study circulating levels of the lysolipids 1-stearoylglycerophosphoinositol and 1-palmitoylglycerophosphoinositol were significantly associated with shortening of LTL. This suggests an increased activity of PLA₂ in advanced biological ageing. This hypothesis is further confirmed by the positive association of the two lysolipids with PLA₂ gene expression in our study.

PLA₂ activity affects, amongst others, the composition and physiology of cell membranes by catalysing the hydrolysis of membrane lipids (Farooqui et al., 2007; Cantin et al., 1992). The integrity of cell membranes and their ability to resist oxidative stress have been shown to be key aspects of biological ageing (Hulbert, 2003). For instance, studies comparing centenarians with younger controls identified alterations of cell membrane composition (Rabini et al., 2002) and particularly depletion of the lysolipid stearylphosphatidylcholine (Collino et al., 2013) as possible causes for longevity.

Figure 4.4 Phospholipase A. Glycerophospholipids are lipid structures that consist of a glycerol backbone (dark blue) and a phosphate group (light blue). They usually bind two fatty acids chains R_1 and R_2 and a polar head group at R_3 . The enzyme *phospholipase A2* (PLA₂) catalyses the first step of their degradation by releasing the R_2 fatty acid from the backbone.



Another regulator of membrane fluidity is the saturation of fatty acids. Both stearic acid and palmitic acid are saturated fatty acids that are known to decrease membrane fluidity, which in turn has been associated with increased susceptibility to disease (Kamada et al., 1983; Aozaki, 1989; Mecocci et al., 1996). In contrast, higher levels of polyunsaturated fatty acid-containing phospholipids were observed in centenarians compared to the elderly (Caprari et al., 1999), suggesting their involvement in retarded biological ageing.

These alterations of membrane composition with biological ageing provide a possible explanation for previously reported association of LTL with e.g. AD (Panossian, 2003).

4.3.2 Gamma-glutamylamino acids

I found two gamma-glutamylamino acids, gamma-glutamyltyrosine and gamma-glutamylphenylalanine, significantly associated with LTL. These are produced by the degradation of glutathione (GSH) and its conjugates catalysed by the enzyme GGT. Gene expression as well as protein abundance of GGT in blood were correlated with both gamma-glutamylamino acids.

The metabolism of gamma-glutamylamino acids by GGT is part of the GSH-cycle and mainly serves the regeneration of the intracellular GSH pool by breaking down extra-cellular GSH conjugates and make its components available for reimport into the cell (Zhang et al., 2005b; Wu et al., 2004; Maher, 2005). GSH is crucial for detoxification of ROS as well as other toxic compounds (Wu et al., 2004; Zhang et al., 2005b; Maher, 2005). Thus, increased GGT activity was proposed as a marker of increased oxidative stress (Pandur et al., 2007; Zhang et al., 2005b). The serum concentration of GGT also is a commonly used clinical marker of liver function (Whitfield, 2001). While the liver produces most of the GSH (Wu et al., 2004), it is most active in kidneys, which absorb GSH for detoxification (Whitfield, 2001; Wu et al., 2004). Accordingly, I found kidney function, measured as eGFR, being highly correlated with both, LTL and gamma-glutamylamino acids.

Summarising, gamma-glutamylamino acids indicate an involvement of increased oxidative stress and worsened kidney and liver function in biological ageing. This highlights the importance of oxidative stress in biological ageing (c.f. section 1.1.1).

4.3.3 4-Vinylphenol sulfate

4-vinylphenol sulfate was previously found to be a marker of tobacco smoking (Manini et al., 2003), which I replicated in this study. Moreover, I found both 4-vinylphenol sulfate and LTL to be correlated with cotinine a metabolite of nicotine and well-established marker for tobacco smoking. Accordingly, higher levels of 4-vinylphenol sulfate were associated with worsened lung function.

Analysis of DNA methylation data from the TwinsUK cohort confirmed the previously published associations of 4-vinylphenol sulfate with the methylation of the RARA gene (Petersen et al., 2014), which I found also associated with LTL and smoking. RARA is a transcription factor that was shown to regulate differentiation and apoptosis (Hu et al., 2014).

My results indicate how smoking might accelerate biological ageing, mediated by changes in metabolism and DNA methylation. Smoking was previously shown to have a profound effect on the GSH metabolism of the lung (Rahman et al., 1999), suggesting increased oxidative stress as a possible link between smoking, metabolism and LTL.

4.4 Conclusion

While I identified five novel markers of LTL, the study has some limitations. First, I analysed data of females only and some of the identified metabolites are known to differ between genders (Krumsiek et al., 2015). In a small pilot of 372 men from the TwinsUK cohort I observed concordant correlations of LTL with the gamma-glutamylamino acids and 4-vinylphenol sulfate. In contrast, neither of the two lysolipids were correlated with LTL in men, suggesting gender-specific changes of fatty acid metabolism with biological ageing. Second, associations did not reach statistical significance in the replication cohort. This can be attributed to smaller sample size, which only allows for 50 % power to detect the observed correlations at nominal significance. However, despite the lack of power, the much higher age, the different geographical location, and genetic background of the replication cohort, all five metabolites showed the same direction in the replication cohort and remained Bonferroni-significant after meta-analysis. Thirdly, integration of the results using multivariate methods, such as GGMs, would permit stronger hypothesis. However, measures of LTL, metabolite levels, gene expression, and clinical phenotyped were available for varying subsets of the TwinsUK cohort with only a small subset of individuals having complete measured for all of them. Thus, multivariate methods that rely on full data-matrices could not be used. Lastly, this association study cannot prove causality. There might be unknown confounding factors, such as age-related diseases. Consequently, the identified metabolites might be markers for ageing rather than causally linked with telomere length. Also, I cannot infer causal direction of the observed correlations.

Nonetheless, this study suggests two mechanisms of biological ageing: First, I observed changes in lipid metabolism, possibly related to membrane composition, strongly associated with LTL and biological ageing. Secondly, I confirmed the involvement of oxidative stress and detoxification of ROS in biological ageing, highlighting the importance of renal health for healthy ageing.

CHAPTER 5

Glycosylation profile of IgG in moderate kidney dysfunction

Here, I analysed the association between renal function and glycosylation of *immunoglobulin G* (IgG). My results highlight the potential involvement of IgG-mediated inflammation in kidney dysfunction.

This work has been done in close cooperation with Clara Barrios, who wrote the original manuscript, while I performed the statistical analyses. The results of this study have been published in *Journal of the American Society of Nephrology* (Barrios, Zierer et al., 2016). The published version is attached in appendix G.3.

Renal function declines irreversibly with age, eventually leading to CKD (see section 1.1.3). Despite a heritability of about 35 %, to date common genetic variation accounts only for 3.2 % of the observed variance (Pattaro et al., 2016), suggesting epigenetic or post-transcriptional mechanism in the development of CKD. One feature of CKD is chronic inflammation, which is associated with faster progression of CKD (Tonelli et al., 2005) and increased risk of comorbidities (Silverstein, 2009).

IgG is a glycoprotein that is part of the immune system, defending the organism against pathogens. Thus, it has a strong inflammatory potential and was shown to be involved in a number of inflammatory diseases, such as arthritis (Parekh et al., 1985; Tomana et al., 1988; Lauc et al., 2013). A main modulator of the inflammatory potential of IgG are attached sugar chains – glycans (Novokmet et al., 2014) (see section 1.2.5).

Animal models highlighted the potential role of IgG glycosylation in the pathophysiological mechanism involved in renal damage. Indeed, studies have shown that modulation of anti-neutrophil cytoplasmic autoantibodies (ANCA) IgG glycosylation reduces its pathogenicity in a mouse model of ANCA-associated glomerulonephritis (Timmeren et al., 2010). Also, IgG Fc γ -receptor deficiency was found to be protective against diabetic nephropathy in mice (Lopez-Parra et al., 2012). Human studies suggest that aberrant glycosylation of the *immunoglobulin A* (IgA) is implicated in the deposit and formation of the immunocomplex IgA-IgG in patients with IgA-Nephropathy (Novak et al., 2005; Novak et al., 2012).

However, despite this evidence for an involvement of IgG-mediated inflammation in kidney disease, possibly due to glycosylation, no studies have investigated this relationship in humans, yet. The aim of this study was to analyse the involvement of IgG glycosylation in renal disease by analysing data from the TwinsUK cohort.

5.1 Methods

I analysed data from 3274 individuals from the TwinsUK cohort with data on IgG glycosylation and renal function available. This included 31 pairs of MZ twins discordant for renal function, defined as difference in eGFR $>15 \text{ mL/min/1.73m}^2$. These were excluded from the discovery cohort and used as independent replication (Table 5.1).

IgG glycosylation was analysed using UPLC by Genos Glycoscience Laboratories (Zagreb, Croatia), measuring the abundance of 24 glycan structures (see section 3.3.2). Additionally, I analysed 52 derived glycan traits that summarise global features of the IgG glycome (Table 3.3). Renal function was measured as eGFR, calculated from the concentration of creatinine in serum using the CKD-EPI equation (section 3.2.3).

To assess the association of IgG glycosylation with renal function I used linear mixed models (section 3.4.2) to regress eGFR against each of the 76 glycan traits, adjusting for age, sex, BMI, and family relatedness by a family-wise random intercept. Additional adjustment for diabetes and hypertension did not change the results. I used the Bonferroni-correction to account for multiple testing (see section 3.4.2.1). Results were then replicated in 31 independent pairs of twins and results were meta-analysed using inverse-variance fixed-effects meta-analysis (see section 3.4.2.2).

Lastly, I used a multivariate LASSO regression, combining four Bonferroni-significant glycans to evaluate the potential of glycosylation markers to predict CKD, defined as eGFR $<60 \text{ mL/min/1.73m}^2$ (see section 3.4.4). The predictive performance of this model was assessed using the ROC statistics evaluated in a ten-fold cross-validation (see section 3.4.6) and compared with the prediction based on age, sex, and BMI only.

Table 5.1 Glycomics marker of renal function: Population characteristics. To identify *immunoglobulin G* (IgG) glycomics markers of renal function, I analysed 3212 individuals from the TwinsUK cohort and replicated the findings in an independent set of 62 monozygotic (MZ) twins discordant for renal function.

	Discovery Population	MZ Discordant Twins
<i>Sample size, n</i>	3212	62
<i>MZ:DZ:singletons</i>	506:1772:934	62:0:0
<i>Age, years</i>	52.67 \pm 14.15	55.45 \pm 12.2
<i>Female, n (%)</i>	3050 (94.9 %)	60 (96.7 %)
<i>BMI, kg/m²</i>	25.95 \pm 4.65	25.64 \pm 5.65
<i>Creatinine,mg/ml</i>	0.83 \pm 0.15	0.75 \pm 0.10
<i>eGFR, mL/min/1.73m²</i>	84.15 \pm 17.02	88.52 \pm 9.91
<i>CKD (eGFR\leq60), n</i>	294 (9.2 %)	1 (1.6 %)
<i>Type II diabetes, n</i>	72 (2.2 %)	4 (6.4 %)
<i>Hypertension, n</i>	705 (21.9 %)	18 (29.0 %)

5.2 Results

Levels of 76 IgG glycosylation markers – 24 directly measured (Figure 3.7) and 52 derived traits (Table 3.3) – were obtained for 3274 individuals from the TwinsUK population, aged between 18 and 97. To validate findings, results were replicated in a group of 31 pairs of MZ twins discordant for renal function, who were excluded from the discovery cohort. The demographic characteristics of the discovery population and the replication set are presented in Table 5.1.

Linear mixed regression models were used to identify 13 glycan traits significantly associated with eGFR, independently of age, sex, BMI, and family relatedness, in the discovery cohort ($p < 6.5 \times 10^{-4} = 0.05/76$ glycan traits); five of them positively associated with eGFR and eight negatively (Table B.1). All 13 glycan traits showed that same direction of effect in an independent set of 62 MZ discordant twins and remained Bonferroni-significant after the meta-analysis (Table 5.2, Figure 5.1). Additionally adjusting the models for menopause did not change the results, suggesting that the observed associations were independent of sexual hormones.

The 14 glycans that were significantly associated with eGFR fell into three glycosylation categories: galactosylation, sialylation and the level of bisecting N-acetylglucosamine (GlcNAc) in the IgG glycome. Firstly, glycans with galactose on both antennae, FA2G2 (GP14: $\beta[95\%CI] = 1.46 [0.85 : 2.06]$, $p = 2.0 \times 10^{-6}$), and the derived trait G2ⁿ ($\beta[95\%CI] = 1.22 [0.63 : 1.81]$, $p = 5.5 \times 10^{-5}$), representing the percentage of digalactosylated structures in neutral IgG glycans, increased in parallel with renal function. On the contrary, I found the agalactosylated glycans A2 (GP2: $\beta[95\%CI] = -0.95 [-1.45 : -0.44]$, $p = 2.5 \times 10^{-4}$; and GP2ⁿ: $\beta[95\%CI] = -0.95 [-1.45 : -0.45]$, $p = 2.2 \times 10^{-4}$) and FA2B (GP6: $\beta[95\%CI] = -1.14 [-1.70 : -0.58]$, $p = 6.8 \times 10^{-5}$) negatively associated with eGFR. The same negative correlation was observed for the derived trait G0ⁿ ($\beta[95\%CI] = -1.16 [-1.75 : -0.57]$, $p = 1.2 \times 10^{-4}$), which combines all agalactosylated structures, indicating a consistent increase in galactosylated and decrease in agalactosylated glycans with eGFR. Secondly, the major sialylated glycan, FA2G2S1 (GP18: $\beta[95\%CI] = 1.44 [0.87 : 2.02]$, $p = 9.5 \times 10^{-7}$), and the percentage of sialylated glycans without bisecting GlcNAc, measured by the ratio FGS / (F + FG + FGS) ($\beta[95\%CI] = 1.00 [0.46 : 1.53]$, $p = 2.9 \times 10^{-4}$), were both associated with increased eGFR. Thirdly, the amount of bisecting GlcNAc in sialylated IgG glycans (FBS^{total} / FS^{total}: $\beta[95\%CI] = -1.04 [-1.55 : -0.52]$, $p = 9.5 \times 10^{-5}$; FBS1 / FS1: $\beta[95\%CI] = -1.10 [-1.62 : -0.58]$, $p = 3.4 \times 10^{-5}$; and FBS1 / (FS1+FBS1): $\beta[95\%CI] = -1.08 [-1.59 : -0.56]$, $p = 4.5 \times 10^{-5}$) and in digalactosylated neutral IgG glycans (FG2ⁿ / (BG2ⁿ + FBG2ⁿ): $\beta[95\%CI] = 0.91 [0.39 : 1.43]$, $p = 5.5 \times 10^{-4}$) were negatively associated with eGFR.

To assess the relevance of these findings for populations with more severe renal damage, associations were validated in eight twins suffering from CKD stages 4 or 5 (average eGFR 24.7 mL/min/1.73m², ranging from 8.0 to 27.3 mL/min/1.73m²), by comparing them to their renal-healthy co-twins (eGFR > 30 mL/min/1.73m²). Despite the lack of statistical power the glycosylation pattern were analogous to those previously observed in the general population (Figure 5.3).

Table 5.2 IgG glycans significantly associated with eGFR. I identified 13 *immunoglobulin G* (IgG) glycan traits significantly associated with estimated glomerular filtration rate (eGFR) in the discovery cohort of 3212 individuals. Findings were replicated in 62 MZ twins discordant for renal function and results were combined using inverse-variance fixed-effects meta-analysis.

Glycan	Description	h^2	Discovery β [95%CI]	p-value	MZ Disc. Twins β [95%CI]	Meta β [95%CI]	p-value
GP18	The percentage of FA2G2S1 glycan in total IgG glycans	0.73	1.48 [0.89: 2.07]	8.6×10^{-7}	0.59 [−2.23: 3.41]	1.44 [0.87: 2.02]	9.5×10^{-7}
GP14	The percentage of FA2G2 glycan in total IgG glycans	0.36	1.46 [0.85: 2.07]	2.9×10^{-6}	1.33 [−1.81: 4.48]	1.46 [0.85: 2.06]	2.0×10^{-6}
GP6 ⁿ	The percentage of FA2B glycan in total neutral IgG glycans (GPn)	0.75	−1.39 [−1.98:−0.80]	3.6×10^{-6}	−0.84 [−3.44: 1.76]	−1.37 [−1.94:−0.79]	3.2×10^{-6}
GP14 ⁿ	The percentage of FA2G2 glycan in total neutral IgG glycans (GPn)	0.47	1.29 [0.68: 1.90]	3.1×10^{-5}	1.99 [−1.70: 5.67]	1.31 [0.71: 1.91]	1.8×10^{-5}
FBS1 / FS1	Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc in total IgG glycans	0.39	−1.12 [−1.65:−0.59]	3.5×10^{-5}	−0.58 [−3.16: 1.99]	−1.10 [−1.62:−0.58]	3.4×10^{-5}
FBS1 / (FS1+FBS1)	The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans in total IgG glycans	0.42	−1.10 [−1.63:−0.57]	4.6×10^{-5}	−0.60 [−3.14: 1.95]	−1.08 [−1.59:−0.56]	4.5×10^{-5}
G2 ⁿ	The percentage of digalactosylated structures in total neutral IgG glycans	0.41	1.20 [0.60: 1.80]	8.8×10^{-5}	1.98 [−1.83: 5.78]	1.22 [0.63: 1.81]	5.5×10^{-5}
GP6	The percentage of FA2B glycan in total IgG glycans	0.75	−1.14 [−1.71:−0.57]	8.9×10^{-5}	−1.01 [−3.78: 1.76]	−1.14 [−1.70:−0.58]	6.8×10^{-5}
FBS ^{total} / FS ^{total}	Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc in total IgG glycans	0.23	−1.07 [−1.60:−0.54]	8.2×10^{-5}	−0.30 [−2.84: 2.23]	−1.04 [−1.55:−0.52]	9.5×10^{-5}
G0 ⁿ	The percentage of agalactosylated structures in total neutral IgG glycans	0.72	−1.16 [−1.76:−0.56]	1.5×10^{-4}	−1.20 [−4.71: 2.31]	−1.16 [−1.75:−0.57]	1.2×10^{-4}
GP2 ⁿ	The percentage of A2 glycan in total neutral IgG glycans (GPn)	0.71	−0.91 [−1.42:−0.40]	5.0×10^{-4}	−2.00 [−4.66: 0.67]	−0.95 [−1.45:−0.45]	2.2×10^{-4}
GP2	The percentage of A2 glycan in total IgG glycans	0.72	−0.90 [−1.42:−0.38]	6.3×10^{-4}	−2.33 [−5.13: 0.47]	−0.95 [−1.45:−0.44]	2.5×10^{-4}
FGS / (F + FG + FGS)	The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans	0.69	1.01 [0.46: 1.56]	3.0×10^{-4}	0.57 [−2.21: 3.35]	1.00 [0.46: 1.53]	2.9×10^{-4}
FG2 ⁿ / (BG2 ⁿ + FBG2 ⁿ)	Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc in total neutral IgG glycans	0.66	0.91 [0.38: 1.44]	7.3×10^{-4}	0.93 [−1.59: 3.44]	0.91 [0.39: 1.43]	5.5×10^{-4}

Figure 5.1 IgG glycosylation markers associated with renal function. (a) Four directly measured glycans and (b) nine derived glycan traits were significantly associated with renal function in the discovery cohort. Green points indicate individuals with impaired renal function ($\text{eGFR} < 60 \text{ mL/min/1.73m}^2$), blue points renal-healthy individuals.

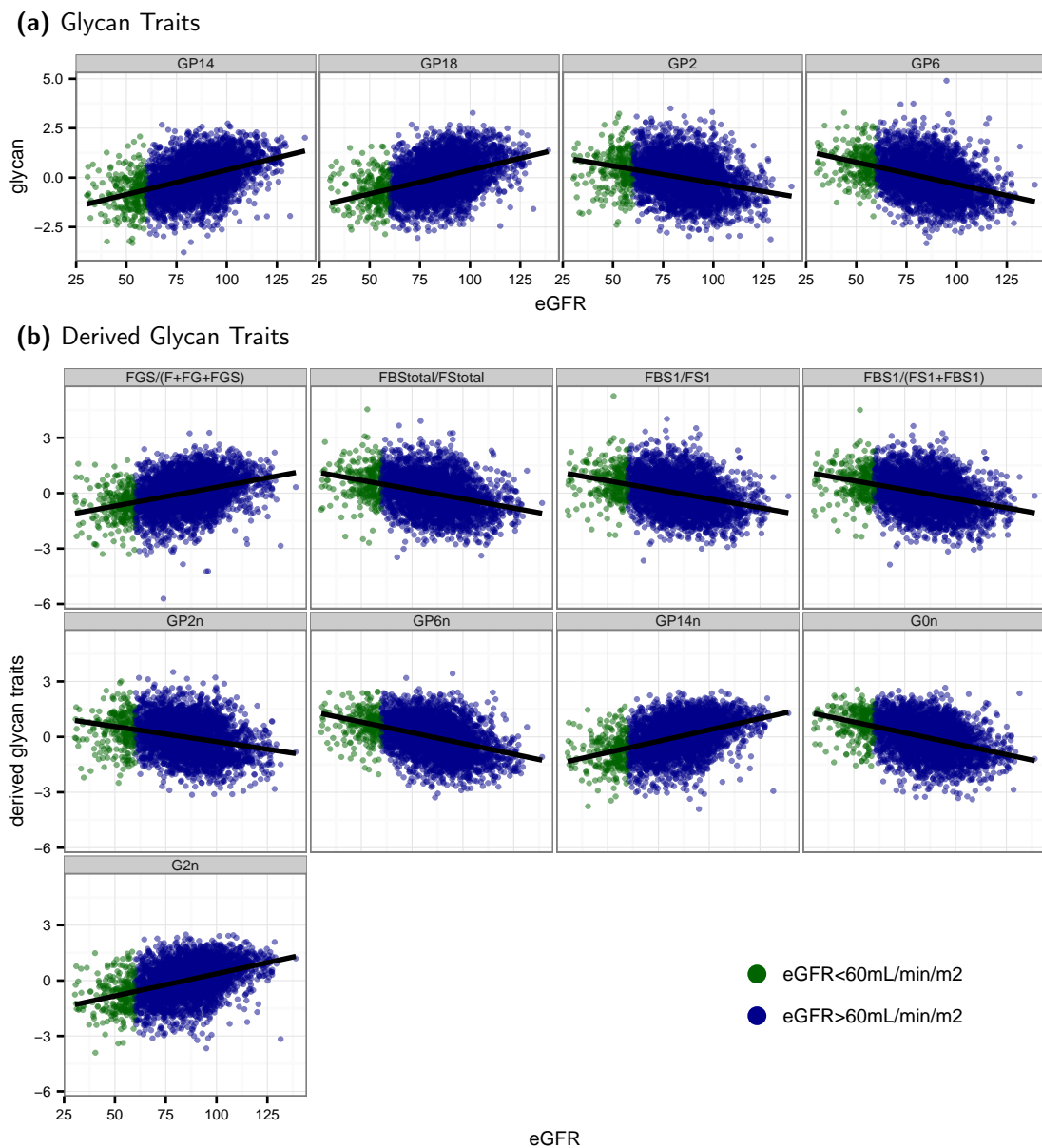
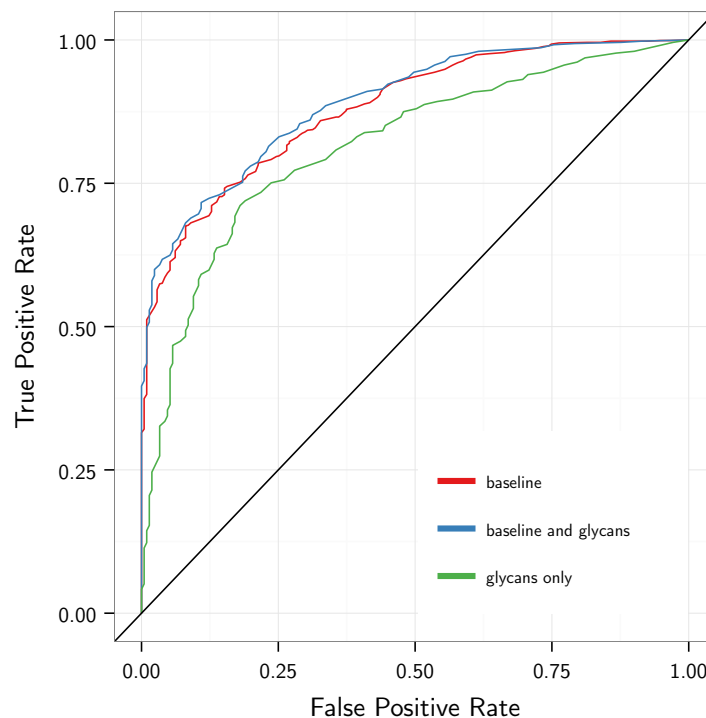


Figure 5.2 Using IgG glycans to predict CKD. Three models were fitted to predict CKD (defined as $\text{eGFR} < 60 \text{ mL/min/1.73m}^2$): The baseline model using age, sex, and BMI as predictors (red), the glycan model using IgG glycosylation traits (green), and the combined model using both (blue). ROC curves were calculated for each of these models using a 10-fold cross-validation.

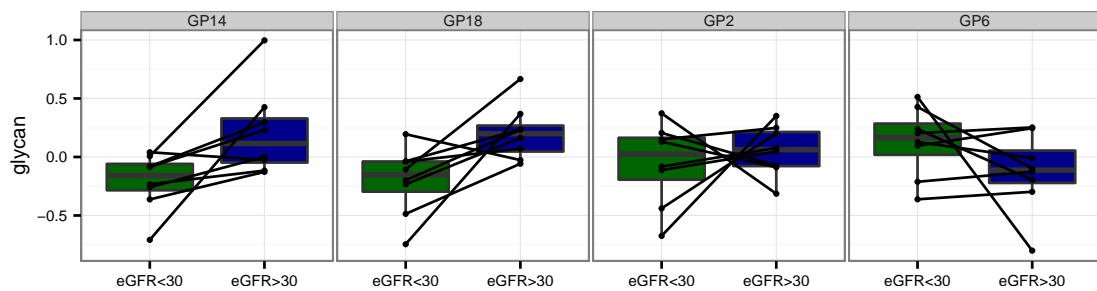


Then, to determine whether the observed changes of glycosylation were specific to IgG or more generally affected multiple circulating proteins, glycosylation profiles of the whole plasma proteome were measured in the same way as the IgG glycosylation. In a subset of 426 individuals none of the whole plasma glycans was significantly associated with eGFR, suggesting that the observed effects are IgG-specific (Table B.2). However, this analysis was under-powered and future studies should investigate the role of global changes in glycosylation.

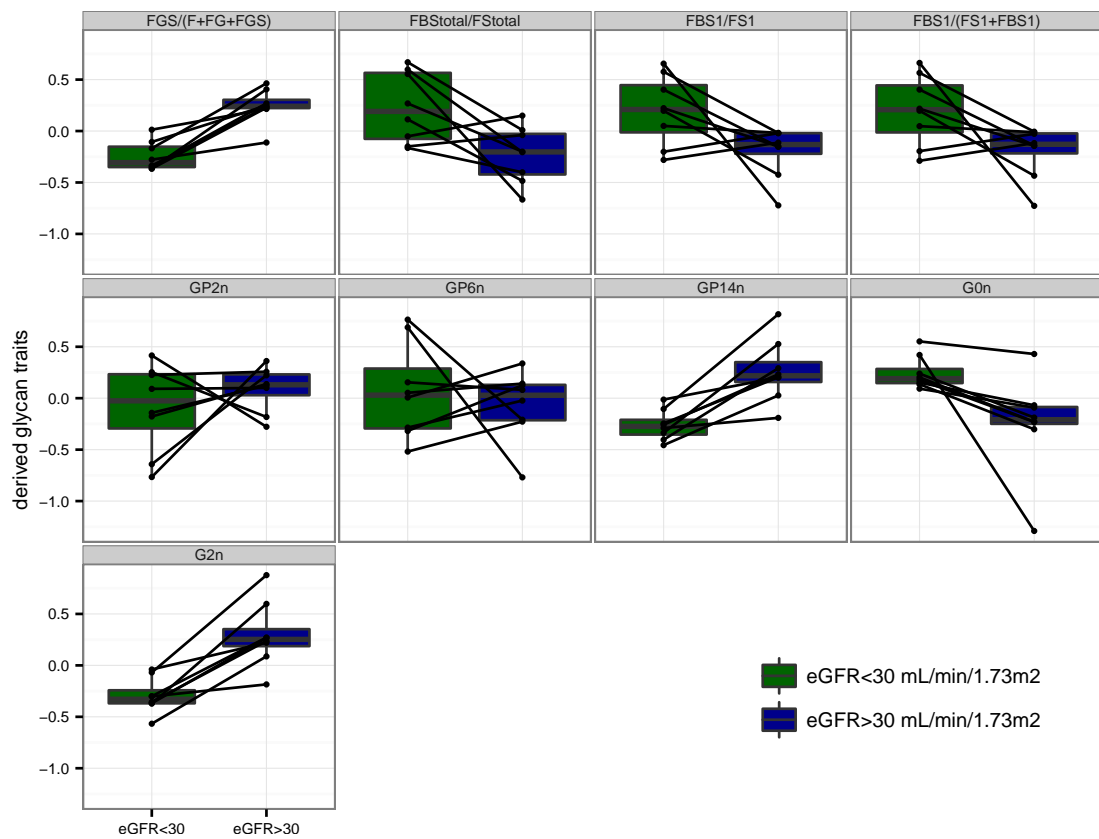
Finally, I assessed the potential of IgG glycosylation to improve the prediction of CKD (defined as $\text{eGFR} < 60 \text{ mL/min/1.73m}^2$) beyond the performance achieved with age, sex and BMI. The predictive ability was assessed using a ROC statistic. The baseline model had an AUC of 0.87 (95 %CI = [0.85 : 0.89]), the model based on the four Bonferroni-significant directly measured glycans had an AUC of 0.81 (95 %CI = [0.78 : 0.84]), and the combined model achieved an AUC of 0.88 (95 %CI = [0.86 : 0.90]) (Figure 5.2).

Figure 5.3 IgG glycosylation in twins discordant for CKD. To assess the transferability of the observed associations from healthy renal function to chronic kidney disease (CKD), associations were confirmed in 8 pairs of twins, one of each twin suffering from severe CKD (stages 4 and 5, $\text{eGFR} < 30 \text{ mL/min/1.73m}^2$).

(a) Glycan Traits



(b) Derived Glycan Traits



5.3 Discussion

This is the first study to investigate the potential role of IgG glycosylation in kidney function. 13 IgG glycan traits were significantly associated with eGFR in the discovery population and were validated in an independent subset of MZ twins discordant for renal function. Moreover, the same pattern was observed in a small, independent sample with severe renal dysfunction. These glycans fall into three glycosylation features, galactosylation, sialylation, and bisecting GlcNAc.

5.3.1 Galactosylation of IgG

In this study, decrease of renal function was associated with a consistent pattern of decreased abundance of galactosylated and increased abundance of agalactosylated IgG. Decreased IgG galactosylation has been found to be associated with chronological and biological ageing (Kristic et al., 2014), as well as several autoimmune and inflammatory diseases (Gornik et al., 2008), including rheumatoid arthritis (Parekh et al., 1985) (c.f. section 1.2.5). Thus, the decrease in galactosylation is not a disease-specific but a rather general feature of biological ageing and the associated immunosuppressive and anti-inflammatory potential of circulating IgG.

The lack of terminal galactose activates the complement cascade and increases the inflammatory potential of IgG, whereas the addition of galactose decreases its inflammatory potential (Baumann et al., 2001; Malhotra et al., 1995). Hence, the IgG galactosylation pattern observed in this study supports theory that complement activation/dysregulation is crucial in renal damage (Cook, 2013). While it is not possible to determine if galactosylation of IgG is a cause or consequence of biological ageing from this study, there is some evidence that it precedes the onset of some diseases, such as rheumatoid arthritis (Rombouts et al., 2015; Ercan et al., 2010). If this holds for CKD, changes in IgG galactosylation might explain a portion of its heritable component, given the high heritability of galactosylated glycans (Menni et al., 2013c).

5.3.2 Sialylation

The major sialylated glycan FA2G2S1 (GP18) and the ratio FGS / (F + FG + FGS), which represents the percentage of sialylated structures without bisecting GlcNAc, appeared to be renal-protective.

The addition of sialic acid to IgG glycans is known to strongly influence its physiologic role, turning it into an anti-inflammatory agent (Anthony et al., 2010; Kaneko et al., 2006). Approximately 50 % of IgG glycans are not sialylated and are hence pro-inflammatory (Huffman et al., 2014). These are, amongst others, thought to be responsible for the immunosuppressive activity of intravenously administered IgG (Anthony et al., 2008).

The amount of sialylation of IgG was previously found to be strongly negatively associated with chronological age (Kristic et al., 2014). The increasing inflammatory potential of IgG due to the loss of sialylation might lead to chronic low-grade inflammation, which has been named

a hallmark of ageing (Franceschi et al., 2014) and was shown to be a risk factor for developing CKD. Changes of IgG sialylation might thus provide an explanation for the continuous decline of renal function with age. However, the causal direction of these associations has to be addressed in future studies. Also, IgG sialylation, like galactosylation, is not a renal-specific marker and has been observed in other diseases, for instance systemic lupus erythematosus (SLE) (Vučković et al., 2015).

5.3.3 Bisecting N-Acetylglucosamine

In this study the eGFR was lower in individuals with bisecting GlcNAc in sialylated and core fucosylated glycans. Accordingly, the proportion of neutral digalactosylated glycans without bisecting GlcNAc ($FG2^n / (BG2^n + FBG2^n)$) was positively associated with eGFR, indicating a deleterious effect of bisecting GlcNAc.

The addition of bisecting GlcNAc was shown to prevent core fucosylation of glycans (Ferrara et al., 2006a), which modulates the antibody-dependent cell-mediated cytotoxicity (ADCC) (Ferrara et al., 2006b). On average, 95 % of the IgG is core fucosylated in a healthy population (Pucic et al., 2011), thus preventing ADCC. Non-fucosylated bisecting glycans, however, bind with much higher affinity to the Fc γ RIIIa and hence initiate antibody-dependent cellular cytotoxicity more efficiently (Masuda et al., 2007; Ferrara et al., 2006b).

Studies using animal models have reported that modifications in the Fc γ receptor can diminish renal damage in autoimmune ANCA-related glomerulonephritis and diabetic nephropathy (Timmeren et al., 2010; Lopez-Parra et al., 2012). However, it remains elusive how modulation of ADCC might affect non-autoimmune decline of renal function.

5.3.4 Genetics of kidney disease

Notably, glycan traits associated with lower eGFR have on average a high heritability (Table 5.2). For instance, agalactosylated IgG glycans that were found to be associated with lower eGFR have a high heritability, ranging from 0.72 to 0.75, whereas the galactosylated glycans GP14 and GP2ⁿ have a low heritability of 0.36 and 0.41, respectively (Menni et al., 2013c).

Several genetic loci influencing IgG glycosylation (Lauc et al., 2013) as well as renal function (Figure 1.3) were previously identified. Interestingly, the gene *ST6 beta-galactoside alpha-2,6-sialyltransferase 1* (ST6GAL1) is associated with both IgA nephropathy (SNP rs7634389, $p = 7.0 \times 10^{-10}$) (Li et al., 2015) and glycosylation of IgG (SNP rs11710456, $p = 6.0 \times 10^{-75}$, $R^2 = 0.46$) (Lauc et al., 2013), particularly the amount of sialylation. ST6GAL1 is a glycosyltransferase that catalyses the sialylation of proteins. Consequently, differential glycosylation of IgG could explain some of the heritability of CKD and mediate the effect of genetic loci on disease development.

5.3.5 Aiding diagnosis of CKD

Although the identified glycans do not predict incident CKD more accurately than clinical parameters, their inclusion in the model slightly improves the predictive performance. However, creatinine is a rather insensitive marker of CKD, particularly of early stages (see section 1.1.3). Thus, identification of biomarkers of early decline of renal function based on creatinine is difficult. Future studies, using more comprehensive renal assessment as well as longitudinal data, might help to assess the predictive power of IgG glycans more reliably.

5.4 Conclusion

This study identified novel markers of renal function, providing new evidence for the importance of inflammatory IgG in the decline of renal function. To minimise false positive discoveries a stringent multiple-testing correction was used and findings were replicated in an independent set of MZ discordant twins. As identical twins share 100 % of their genetic variation and are matched perfectly for age, gender, and social class, they facilitated the validation of the association between IgG and renal function independently of genetic influences.

The study has some limitations. First, there is a female predominance in the study sample (95 % of the study population is female). Secondly, the study participants were recruited as volunteers and consequently have for instance a low prevalence of diabetes. Thus, associations might not be directly transferable to diabetic populations. Also, they are all twins and results might differ for singletons. Thirdly, I analysed cross-sectional data only. Longitudinal studies and particularly prospective analyses will have to confirm these initial results, not to speak of causal relationships, which cannot be addressed with the available data. Finally, the experimental procedure cannot distinguish Fc from Fab glycans, however, the observed correlations are likely to be driven by the much more abundant Fc glycans. To further support this hypothesis, I analysed a pilot dataset of 96 age-matched individuals from the extremes of the eGFR-distribution, for whom Fc glycans were additionally analysed using nano-liquid chromatography MS/MS (Huffman et al., 2014). Effect directions were consistent for all those glycans except one (Table B.3), suggesting that the observed associations of renal function with IgG glycosylation indeed mainly represent changes in Fc glycans.

These findings highlight the promising role of glycomics in renal studies. While this is only an initial finding, future studies on clinical and longitudinal dataset need to assess the potential use of IgG glycosylation for timely diagnosis and treatment of CKD. Cohorts with more detailed information on renal function, including proteinuria, and different types of CKD, such as autoimmune CKD, will help to differentiate CKD-specific glycosylation patterns from more general changes due to biological ageing.

CHAPTER 6

Metabolic markers of renal disease in type 2 diabetics compared to non-diabetics

The aim of this chapter was to identify metabolic markers of renal function in four cohorts, including 926 type 2 diabetes patients and 4838 non-diabetics. While previous studies investigated associations of metabolites with renal function in either diabetic or non-diabetic cohorts, no study has yet addressed the differences between the two. Also, the targeted NMR platform used here can – in contrast to previously used MS-based platforms – distinguish lipid subclasses at a high resolution.

Chronic kidney disease (CKD) is a major public health problem that affects more than 10 % of the Western populations and more than 30 % of the elderly (O’Callaghan et al., 2011) (see section 1.1.3). Diabetes is one of the strongest risk factors for developing kidney disease (Jha et al., 2013) and the leading cause of ESRD, accounting for roughly half of the patients progressing to renal failure (United States Renal Data System, 2016). Due to the continuous increase in the prevalence of T2D (Seidell, 2000; World Health Organization, 2016), because of the global obesity epidemic, the prevalence of diabetic nephropathy is increasing accordingly. Despite the efforts in improving diabetes management, the rate of ESRD among diabetics has been reduced to a much lesser extent than other complications of diabetes (Gregg et al., 2014). This is partly because the commonly used clinical biomarkers of renal function, namely creatinine and albuminuria, are insensitive and unspecific (see section 1.1.3). Thus, it is crucial to identify better biomarkers to facilitate early diagnosis and treatment of renal disease.

In recent years, several studies investigated metabolic profiles associated with renal function (Zhang et al., 2015; Barrios et al., 2016) in the general population (Sekula et al., 2016) and in type 1 diabetes patients (Mäkinen et al., 2013; Mäkinen et al., 2012) to identify biomarkers predicting disease progression (Mäkinen et al., 2012) as well as complications of renal disease and increased mortality (Mäkinen et al., 2013). While some studies investigated metabolic markers of (type 2) diabetic kidney disease, these studies were mostly based on very small samples sizes and lacked external replication (Darshi et al., 2016).

Here, I investigated metabolic signatures of renal damage in the largest study of its kind, combining four European cohorts, to identify differences between metabolic profiles associated with diabetic and non-diabetic nephropathy. Using a targeted NMR-based platform facilitates

the in-depth analysis of lipoprotein composition, which might be related to the increased risk of cardiovascular events in CKD patients (Shen et al., 2016; Kon et al., 2015). Additionally, longitudinal changes of metabolite levels and renal function and associations with other microvascular complications of diabetes were analysed to gain insights in potential mechanisms of the cross-sectional associations.

6.1 Methods

6.1.1 Study populations

Markers of diabetic nephropathy were analysed in the GenoDiabMar cohort, while the TwinsUK, KORA, and YoungFinns studies represented the general population. Diabetics from the TwinsUK and KORA cohorts were analysed separately and used as replication sets for the GenoDiabMar cohort.

6.1.1.1 GenoDiabMar

Here, I analysed 655 diabetic individuals of the GenoDiabMar cohort (section 3.1.2.1) with NMR metabolomics data, measured by Brainshake Ltd., available. Follow-up measurements of clinical and analytical data were available for a subset of 326 individuals.

6.1.1.2 TwinsUK

1279 individuals of the TwinsUK cohort (section 3.1.1) with NMR metabolomics data and enzymatic measurements of creatinine were included. The study sample included 111 T2D individuals, which were analysed separately. Also, longitudinal measures of creatinine and NMR metabolomics were available for 740 individuals, 363 of them examined at three visits, with an average follow-up time of (11.8 ± 4.7) years.

6.1.1.3 KORA

For the present analysis, 1784 participants of the KORA study (section 3.1.2) with NMR metabolomics and clinical creatinine measurements available were included. 160 individuals, who have been diagnosed with T2D, were analysed separately. Follow-up measurements of clinical creatinine were available for a subset of 1185 individuals, 7 years after the baseline visit.

6.1.1.4 YoungFinns

Lastly, I analysed data from the 21, 27, and 30 years follow-ups of the YoungFinns cohort (section 3.1.4). Longitudinal measurements of creatinine and NMR metabolomics at all three time points were available for 1770 of the 2046 individuals.

6.1.2 Metabolomic profiling

Metabolomics data was measured for all cohorts by Brainshake Ltd. using ^1H NMR spectroscopy, as described in section 3.3.1.2. The platform quantifies absolute concentrations of 144 metabolites, including 98 lipoproteins from 14 lipoprotein subclasses and 9 amino acids. Moreover, it provides measures of average particle sizes for very low density lipoprotein (VLDL), low density lipoprotein (LDL), intermediate density lipoprotein (IDL), and HDL particles as well as a semi-quantitative measure of albumin concentration. Additionally, 80 ratios of metabolites describing the composition of lipoprotein particles were analysed, totalling to 227 metabolic traits (Table C.3).

All metabolic phenotypes were log-transformed. To account for zero values a pseudo-count of 1 was added to all measurements prior to transformation. Outliers, differing more than 6 standard deviations (SDs) from the population mean, were excluded. Prior to analysis all measurements were scaled to mean zero and SD of 1 to facilitate comparisons across cohorts.

6.1.3 Phenotypes

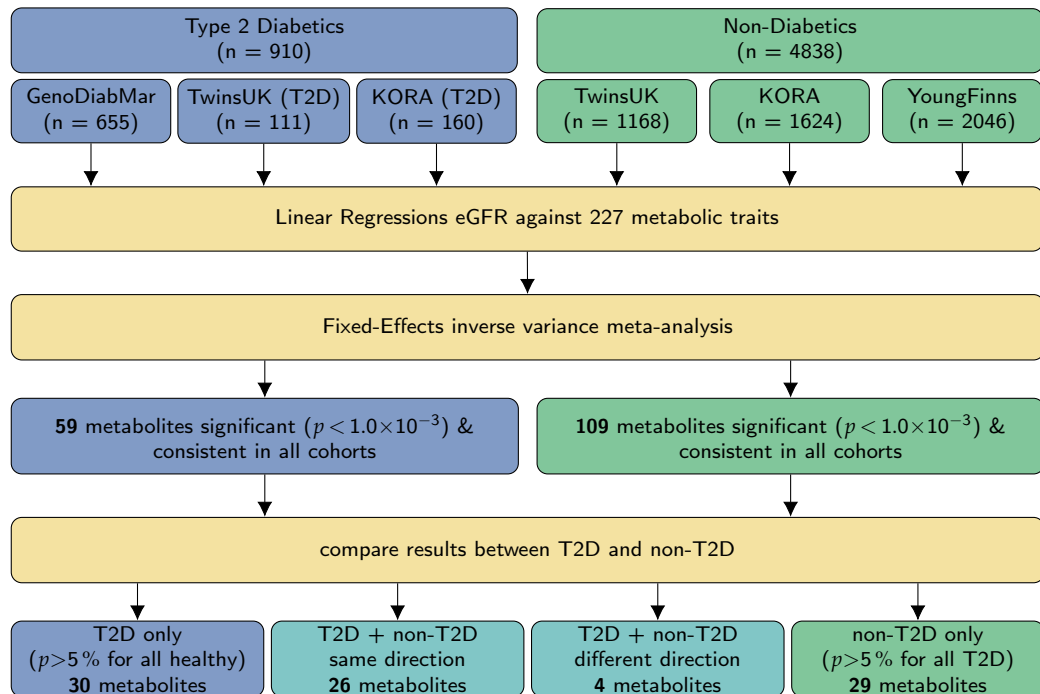
For all cohorts, renal function was measured as eGFR from enzymatic measurements of serum creatinine (see section 3.2.3). Age, gender, and BMI, calculated as mass in kg over the square of height in meters (section 3.2.1), were additionally collected as confounding factors. Analyses were stratified by type 2 diabetes status. All participants of the GenoDiabMar cohort were diagnosed with T2D by a doctor and receive treatment. In TwinsUK and KORA 111 and 160 participants, respectively, reported to be diabetic and were, thus, analysed separately.

I additionally analysed albuminuria and diabetic retinopathy, another microvascular complication of diabetes, in the GenoDiabMar cohort.

6.1.4 Statistical analysis

6.1.4.1 Cross-sectional analysis

First, I assessed the associations between metabolic profiles and renal function in each cohort by fitting linear regression models for all metabolic traits with eGFR as outcome, adjusting for age, gender, and BMI (section 3.4.2). Regression models were additionally adjusted for family structure by adding a family-wise random intercept in the TwinsUK cohort. As only aggregate data was available for some of the cohorts, each cohort was analysed separately and results were meta-analysed separately for T2D patients and non-diabetic cohorts using inverse variance fixed effect meta-analysis implemented in the R package *meta* (section 3.4.2.2). Only metabolites with consistent effect directions in all cohorts were further considered (Figure 6.1).

Figure 6.1 Flowchart illustrating the identification of metabolic markers of diabetic and non-diabetic renal disease.**Table 6.1 Metabolic markers of renal function: population characteristics.** I analysed four different cohorts. GenoDiabMar consists of type 2 diabetes (T2D) patients only, and YoungFinns has no T2D patients. TwinsUK and KORA were stratified by diabetes status and analysed separately, resulting in three diabetic and three non-diabetic groups. Population averages and standard deviations are given for all continuous phenotypes. Grades of chronic kidney disease (CKD) were defined according to KDIGO (Levey et al., 2005) recommendations (c.f. Table 1.1)

	Diabetics			Non - Diabetics		
	GenoDiab Mar	TwinsUK (diabetics)	Kora (diabetics)	TwinsUK (non T2D)	Kora (non T2D)	Young Finns
<i>N</i>	655	111	160	1168	1624	2046
<i>Zygosity (MZ/DZ/Single)</i>		30/4/77		466/546/156		
<i>Age, years</i>	69.70 (±9.32)	68.64 (±8.38)	66.74 (±7.44)	64.83 (±7.91)	60.30 (±8.83)	41.88 (±5.00)
<i>Gender (female)</i>	256 (39.1 %)	105 (94.6 %)	71 (44.4 %)	1118 (95.7 %)	845 (52.0 %)	1115 (54.5 %)
<i>BMI, kg/m²</i>	30.32 (±5.05)	29.33 (±5.55)	31.48 (±5.53)	26.05 (±4.61)	27.82 (±4.58)	26.54 (±5.05)
<i>eGFR, mL/min/1.73m²</i>	58.64 (±28.83)	75.80 (±17.64)	76.59 (±18.15)	79.87 (±14.53)	87.80 (±15.38)	94.75 (±12.53)
<i>CKD, renal healthy / grades 2-5</i>	116/213/202/72/52	20/76/12/3/0	39/95/24/1/1	322/726/118/2/0	795/752/73/2/2	1339/699/7/1/0

As the lipid subclasses are very highly correlated, I estimated the number of effective independent variables M_{eff} for each cohort using the eigenvalues of the correlation matrix (Li et al., 2005), as

$$M_{\text{eff}} = \sum_{i=1}^M f(|\lambda_i|)$$

$$f(x) = I(x \geq 1) + (x - \lfloor x \rfloor)$$

where λ_i is the i -th of M eigenvectors and $I(x)$ is the indicator functions returning 1 if x is true. The highest estimate for the number of independent variables was 49.5 (GenoDiabMar cohort). Thus, all analysed were adjusted for 50 independent tests using Bonferroni correction ($p < 1.0 \times 10^{-3} = 0.05/50$).

6.1.4.2 Longitudinal analysis

Longitudinal follow-ups for both enzymatic measures of creatinine and NMR metabolomics were available for TwinsUK and YoungFinns cohorts. I calculated the longitudinal trajectories of eGFR and each of the metabolites over time by fitting linear mixed models, regressing each of these variables against the linear time since baseline as per-individual random effect using the `lmer` function implemented in the R package `lme4` (version 1.1) (Bates et al., 2015). The estimates of per-individual random slopes of these models were used as estimates of the longitudinal trajectories. Then, I regressed the trajectory of eGFR against each of the metabolite trajectories to analyse longitudinally correlations.

A follow-up for clinical creatinine was available for subsets of 326 individuals from the GenoDiabMar cohort and 1185 individuals from the KORA cohort. I evaluated the potential of metabolite levels as diagnostic tool by predicting the eGFR at follow-up using metabolite levels at baseline, correcting for baseline eGFR.

6.2 Results

Associations of circulating metabolites were analysed in 5764 individuals from four independent European cohorts, including 926 T2D patients. The demographic characteristics of all cohorts are presented in Table 6.1. The same targeted NMR platform was used in all cohorts to quantify 227 metabolic traits. Analyses were stratified by diabetes status for the TwinsUK and KORA cohorts, while the GenoDiabMar had T2D patients and the YoungFinns study non-diabetic controls, only. Results from all cohorts were meta-analysed for diabetic and non-diabetics separately (Figure 6.1).

6.2.1 Common markers of renal function

Generally, associations were largely concordant between diabetics and non-diabetics groups (Figure 6.2). 26 of metabolites showed consistent association with eGFR in diabetics and non-diabetics (Table 6.3a).

The strongest associations with eGFR were observed for phenylalanine ($\beta[95\%CI] = -7.92 [-9.27 : -6.57]$, $p = 1.2 \times 10^{-30}$ for T2D; $\beta[95\%CI] = -1.69 [-2.07 : -1.32]$, $p = 1.1 \times 10^{-18}$ for non-T2D), glycine ($\beta[95\%CI] = -8.37 [-9.73 : -7.02]$, $p = 7.3 \times 10^{-34}$ for T2D; $\beta[95\%CI] = -1.29 [-1.66 : -0.92]$, $p = 6.3 \times 10^{-12}$ for non-T2D), and citrate ($\beta[95\%CI] = -3.34 [-4.78 : -1.90]$, $p = 5.7 \times 10^{-6}$ for T2D; $\beta[95\%CI] = -1.82 [-2.18 : -1.47]$, $p = 7.1 \times 10^{-24}$ for non-T2D). Citrate also was significantly correlated with longitudinal change of eGFR ($p = 1.5 \times 10^{-6}$) in TwinsUK, however did not predict eGFR at follow-up in KORA (Table C.2).

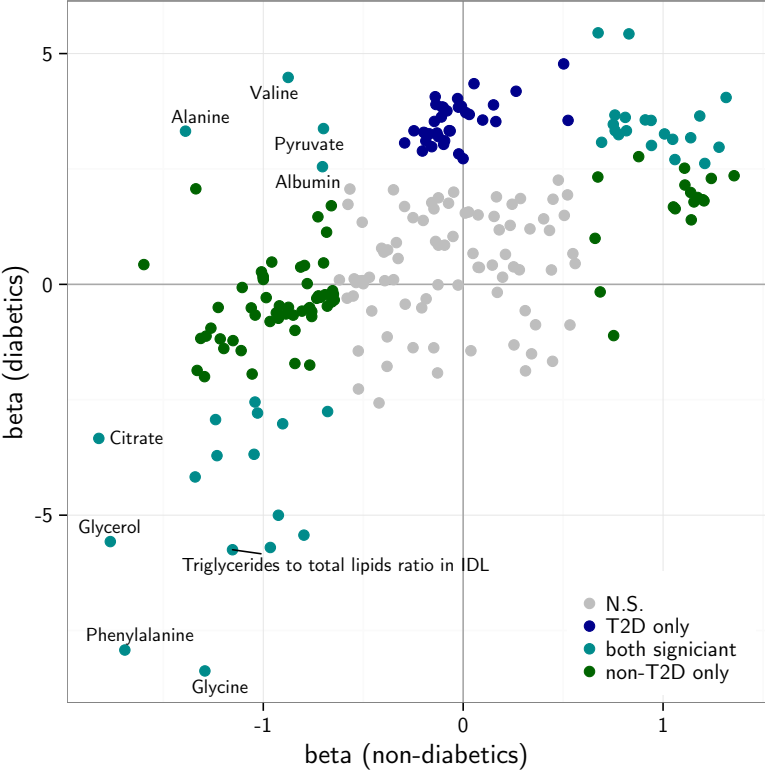
Generally, HDL subclasses (e.g. the concentration of medium HDL particles: $\beta[95\%CI] = 3.33 [1.89 : 4.76]$, $p = 5.4 \times 10^{-6}$ for T2D; $\beta[95\%CI] = 0.76 [0.40 : 1.12]$, $p = 3.3 \times 10^{-5}$ for non-T2D) and cholesterol esters (e.g. cholesterol esters to total lipids ratio in IDL: $\beta[95\%CI] = 5.45 [4.10 : 6.79]$, $p = 2.1 \times 10^{-15}$ for T2D; $\beta[95\%CI] = 0.67 [0.32 : 1.03]$, $p = 1.7 \times 10^{-4}$ for non-T2D) were positively associated with eGFR. Both were also associated with decreased albuminuria ($\beta[95\%CI] = -0.43 [-0.68 : -0.16]$, $p = 9.4 \times 10^{-4}$; and $\beta[95\%CI] = -0.69 [-0.94 : -0.43]$, $p = 1.6 \times 10^{-7}$) as well as decreased risk of diabetic retinopathy ($OR[95\%CI] = 0.76 [0.62 : 0.93]$, $p = 6.8 \times 10^{-3}$; and $OR[95\%CI] = 0.64 [0.53 : 0.78]$, $p = 6.6 \times 10^{-6}$) in the GenoDiabMar cohort (Table C.1).

In contrast, triglycerides were consistently associated with worse renal function in diabetics and non-diabetics (e.g. triglycerides to total lipids ratio in large LDL: $\beta[95\%CI] = -5.70 [-7.04 : -4.36]$, $p = 9.1 \times 10^{-17}$ for T2D; $\beta[95\%CI] = -0.96 [-1.32 : -0.61]$, $p = 7.7 \times 10^{-8}$ for non-T2D), as well as increased albuminuria ($\beta[95\%CI] = 0.68 [0.42 : 0.94]$, $p = 5.0 \times 10^{-7}$) and increased risk for diabetic retinopathy ($OR[95\%CI] = 1.36 [1.13 : 1.63]$, $p = 9.1 \times 10^{-4}$) (Table C.1).

Table 6.2 Metabolites differently associated with eGFR in diabetics and non-diabetics. Four metabolites were significantly associated with estimated glomerular filtration rate (eGFR) in opposite directions in diabetics and non-diabetics, though not consistently across cohorts.

Metabolite	T2D				non-diabetics			
	N	signs	β [95 %CI]	p-value	N	signs	β [95 %CI]	p-value
Alanine	914	++-	3.32 [1.92: 4.72]	3.5×10^{-6}	4715	---	-1.39 [-1.75:-1.03]	2.4×10^{-14}
Valine	914	++-	4.48 [3.08: 5.88]	3.5×10^{-10}	4708	---	-0.88 [-1.27:-0.48]	1.4×10^{-5}
Albumin	926	+-	2.55 [1.16: 3.93]	3.1×10^{-4}	4816	---	-0.70 [-1.06:-0.35]	9.4×10^{-5}
Pyruvate	908	+++	3.37 [1.93: 4.82]	4.7×10^{-6}	4704	+-	-0.70 [-1.06:-0.34]	1.3×10^{-4}

Figure 6.2 Comparison of metabolite associations with renal function between type 2 diabetics and non-diabetics. I compared associations of metabolites with eGFR between diabetics and non-diabetic cohorts. Here, I show effect sizes from both meta-analyses, coloured according to significance level in diabetics (blue), non-diabetics (green) and both (cyan). Non-significant associations are shown in grey. Details are shown in Table 6.3.



6.2.2 Markers of renal function in diabetics

In the three T2D cohorts, 59 metabolites were significantly and consistently associated with eGFR after meta-analysis and adjusting for multiple testing using Bonferroni correction. 30 of these metabolites were associated with eGFR only in diabetics with no evidence of an association in non-diabetics ($p > 0.05$) (Table 6.3b). 22 of those were concentration of different LDL subclasses, particularly cholesterol and cholesterol esters, and 6 were IDL subclasses that followed a positive association with eGFR in diabetics. Also, esterified cholesterol ($\beta[95\%CI] = 4.35 [2.96 : 5.74]$, $p = 9.3 \times 10^{-10}$), total cholesterol ($\beta[95\%CI] = 3.86 [2.29 : 5.07]$, $p = 2.0 \times 10^{-7}$), total cholesterol in IDL ($\beta[95\%CI] = 3.90 [2.53 : 5.36]$, $p = 2.1 \times 10^{-8}$), and total cholesterol in LDL ($\beta[95\%CI] = 3.84 [2.48 : 5.20]$, $p = 3.4 \times 10^{-8}$) were all positively associated with eGFR. However, neither of those showed significant correlations with albuminuria or diabetic retinopathy.

6.2.3 Markers of renal function in non-diabetics

In the three non-diabetic cohorts 109 metabolites were associated with eGFR (Table C.3) after meta-analysis and adjustment for multiple testing. 29 of these were associated with renal function only in non-diabetics, showing no evidence of correlation in any of the diabetic cohorts ($p > 0.05$) (Table 6.3c). Particularly total cholesterol (6), cholesterol esters (5) and free cholesterol (4) levels in VLDL particles of different sizes were consistently negatively associated with eGFR in diabetics, but not non-diabetics.

In contrast, Phospholipids to total lipids ratio in very large HDL were positively associated with eGFR cross-sectionally ($\beta[95\%CI] = 1.06 [0.65 : 1.47]$, $p = 4.8 \times 10^{-7}$) and longitudinally ($\beta[95\%CI] = 0.04 [0.02 : 0.06]$, $p = 6.7 \times 10^{-6}$) in TwinsUK. Also, the ratio predicted longitudinal change of eGFR independently of baseline eGFR in KORA ($\beta[95\%CI] = 0.04 [0.01 : 0.07]$, $p = 3.7 \times 10^{-3}$).

6.2.4 Metabolites differently associated with eGFR

Finally, four metabolites were associated with eGFR in opposite directions in diabetic and non-diabetic cohorts (Table 6.2): The amino acids alanine ($\beta[95\%CI] = 3.32 [1.92 : 4.72]$, $p = 3.5 \times 10^{-6}$ in T2D and $\beta[95\%CI] = -1.39 [-1.75 : -1.03]$, $p = 2.4 \times 10^{-14}$ in non-T2D) and valine ($\beta[95\%CI] = 4.48 [3.08 : 5.88]$, $p = 3.5 \times 10^{-10}$ in T2D and $\beta[95\%CI] = -0.88 [-1.27 : -0.48]$, $p = 1.4 \times 10^{-5}$ in non-T2D), pyruvate ($\beta[95\%CI] = 3.37 [1.93 : 4.82]$, $p = 4.7 \times 10^{-6}$ in T2D and $\beta[95\%CI] = -0.70 [-1.06 : -0.34]$, $p = 1.3 \times 10^{-4}$ in non-T2D), and albumin ($\beta[95\%CI] = 2.55 [1.16 : 3.93]$, $p = 3.1 \times 10^{-4}$ in T2D and $\beta[95\%CI] = -0.70 [-1.06 : -0.35]$, $p = 9.4 \times 10^{-5}$ in non-T2D). All of those were negatively associated with eGFR in the non-diabetics but positively in the diabetics, though not consistently across cohorts. Albumin also predicted the longitudinal change of eGFR ($\beta[95\%CI] = 0.35 [0.12 : 0.58]$, $p = 3.4 \times 10^{-3}$) (Table C.2) in the GenoDiabMar cohort.

Figure 6.3 Metabolites associated with eGFR in diabetic and non-diabetic cohorts. 227 metabolites were associated with the estimated glomerular filtration rate (eGFR) three cohorts of type 2 diabetics and three cohorts of non-diabetics. Fixed effects inverse meta-analysis was used to combine results for both groups. Associations with $p < 1.0 \times 10^{-3}$ were considered significant. For detailed list of results and full metabolites names see Table C.3.

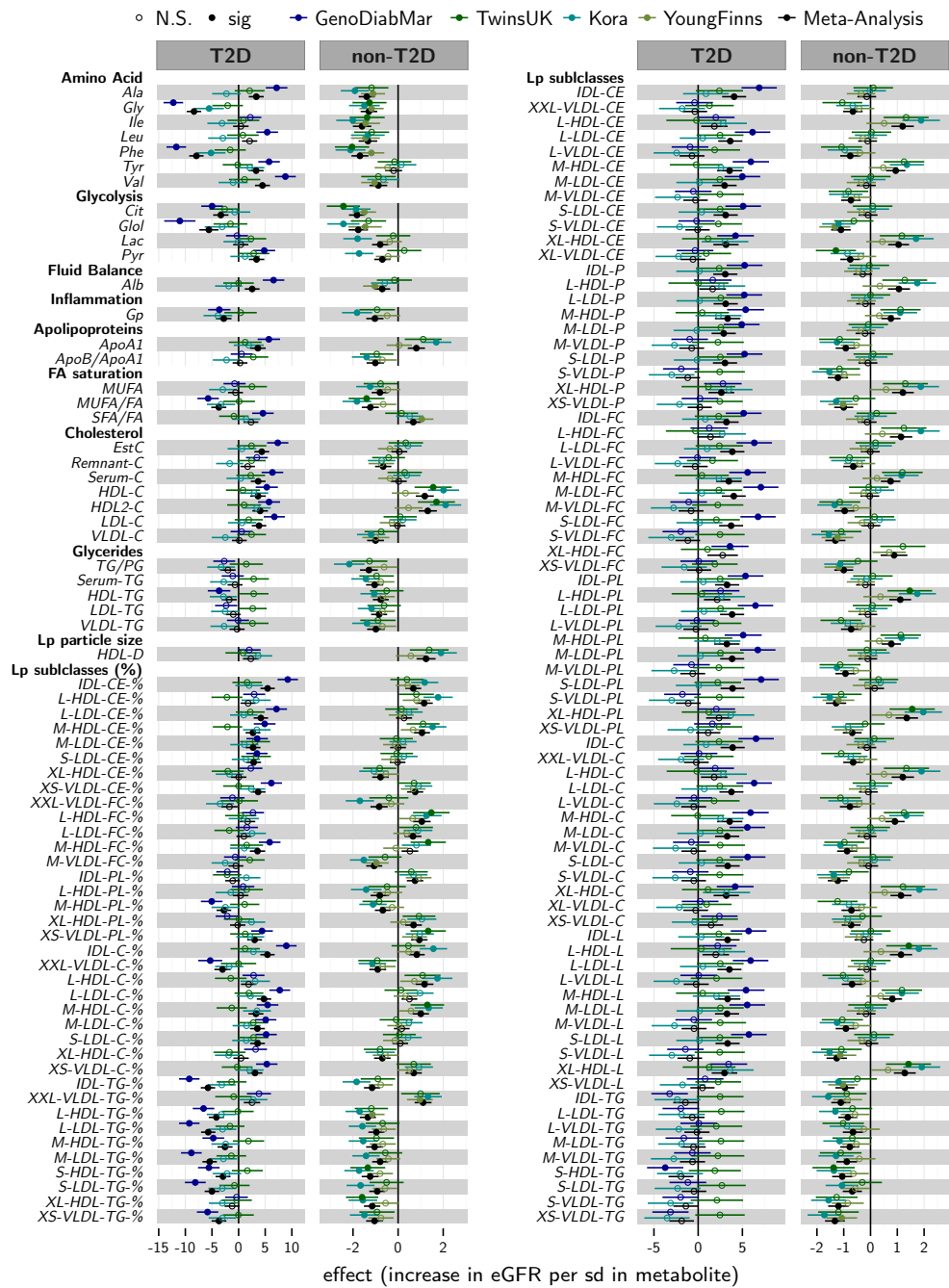


Table 6.3 Metabolites associated with renal function. 227 metabolites were analysed in two groups of cohorts, diabetics and non-diabetics.

Here metabolites are categorised in (a) metabolites consistently associated with estimated glomerular filtration rate (eGFR) in diabetics and non-diabetics , (b) metabolites associated with eGFR only in diabetics, and (c) metabolites associated with eGFR in non-diabetics only.

(a) Metabolites consistently associated with eGFR in diabetics and non-diabetics

Metabolite	Group	T2D				non-diabetics			
		N	signs	β [95 %CI]	p-value	N	signs	β [95 %CI]	p-value
Glycine	Amino Acid	887	- - -	-8.37[-9.73;-7.02]	7.3×10^{-34}	4632	- - -	-1.29[-1.66;-0.92]	6.3×10^{-12}
Phenylalanine	Amino Acid	905	- - -	-7.92[-9.27;-6.57]	1.2×10^{-30}	4716	- - -	-1.69[-2.07;-1.32]	1.1×10^{-18}
Apolipoprotein A-I	Apolipoproteins	926	+ + +	3.62[2.14: 5.10]	1.6×10^{-6}	4817	+ + +	0.81[0.43: 1.19]	3.1×10^{-5}
Total cholesterol in HDL	Cholesterol	925	+ + +	3.65[2.19: 5.11]	9.6×10^{-7}	4817	+ + +	1.18[0.78: 1.58]	6.6×10^{-9}
Total cholesterol in HDL2	Cholesterol	924	+ + +	4.05[2.59: 5.50]	4.9×10^{-8}	4817	+ + +	1.32[0.91: 1.72]	1.5×10^{-10}
Citrate	Glycolysis and Gluconeogenesis	913	- - -	-3.34[-4.78;-1.90]	5.7×10^{-6}	4705	- - -	-1.82[-2.18;-1.47]	7.1×10^{-24}
Glycerol	Glycolysis and Gluconeogenesis	551	- - -	-5.57[-7.37;-3.77]	1.2×10^{-9}	3695	- - -	-1.77[-2.19;-1.34]	7.5×10^{-16}
Cholesterol esters in very large HDL	Lipoprotein subclasses	926	+ + +	3.14[1.71: 4.57]	1.6×10^{-5}	4817	+ + +	1.05[0.66: 1.44]	1.4×10^{-7}
Cholesterol esters to total lipids ratio in IDL	Lipoprotein subclasses	917	+ + +	5.45[4.10: 6.79]	2.1×10^{-15}	4807	+ + +	0.67[0.32: 1.03]	1.7×10^{-4}
Cholesterol esters to total lipids ratio in very small VLDL	Lipoprotein subclasses	920	+ + +	3.67[2.28: 5.06]	2.3×10^{-7}	4808	+ + +	0.76[0.40: 1.12]	3.1×10^{-5}
Concentration of medium HDL particles	Lipoprotein subclasses	926	+ + +	3.33[1.89: 4.76]	5.4×10^{-6}	4817	+ + +	0.76[0.40: 1.12]	3.3×10^{-5}
Concentration of very large HDL particles	Lipoprotein subclasses	926	+ + +	2.62[1.17: 4.07]	4.1×10^{-4}	4817	+ + +	1.21[0.80: 1.62]	5.9×10^{-9}
Free cholesterol in medium HDL	Lipoprotein subclasses	926	+ + +	3.47[2.03: 4.90]	2.2×10^{-6}	4817	+ + +	0.75[0.39: 1.11]	5.3×10^{-5}
Phospholipids in medium HDL	Lipoprotein subclasses	926	+ + +	3.24[1.80: 4.68]	1.0×10^{-5}	4817	+ + +	0.78[0.42: 1.14]	2.4×10^{-5}
Phospholipids to total lipids ratio in very small VLDL	Lipoprotein subclasses	923	+ + +	3.01[1.61: 4.40]	2.3×10^{-5}	4808	+ + +	0.94[0.58: 1.31]	4.3×10^{-7}
Total cholesterol in very large HDL	Lipoprotein subclasses	926	+ + +	3.18[1.74: 4.61]	1.4×10^{-5}	4817	+ + +	1.14[0.74: 1.53]	1.6×10^{-8}
Total cholesterol to total lipids ratio in IDL	Lipoprotein subclasses	916	+ + +	5.43[4.08: 6.77]	2.6×10^{-15}	4805	+ + +	0.83[0.48: 1.18]	3.6×10^{-6}
Total cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	Lipoprotein subclasses	739	- - -	-3.02[-4.55;-1.49]	1.1×10^{-4}	4005	- - -	-0.90[-1.28;-0.53]	2.2×10^{-6}
Total lipids in medium HDL	Lipoprotein subclasses	923	+ + +	3.33[1.88: 4.77]	6.3×10^{-6}	4813	+ + +	0.82[0.46: 1.18]	8.8×10^{-6}
Total lipids in very large HDL	Lipoprotein subclasses	926	+ + +	2.97[1.52: 4.42]	5.7×10^{-5}	4817	+ + +	1.28[0.87: 1.69]	8.4×10^{-10}
Triglycerides to total lipids ratio in IDL	Lipoprotein subclasses	923	- - -	-5.75[-7.10;-4.40]	8.2×10^{-17}	4811	- - -	-1.15[-1.51;-0.80]	2.4×10^{-10}
Triglycerides to total lipids ratio in large HDL	Lipoprotein subclasses	886	- - -	-4.17[-5.60;-2.75]	9.5×10^{-9}	4612	- - -	-1.34[-1.70;-0.98]	1.7×10^{-13}
Triglycerides to total lipids ratio in large LDL	Lipoprotein subclasses	923	- - -	-5.70[-7.04;-4.36]	9.1×10^{-17}	4813	- - -	-0.96[-1.32;-0.61]	7.7×10^{-8}
Triglycerides to total lipids ratio in medium LDL	Lipoprotein subclasses	913	- - -	-5.43[-6.79;-4.08]	4.1×10^{-15}	4808	- - -	-0.80[-1.15;-0.44]	9.3×10^{-6}
Triglycerides to total lipids ratio in small LDL	Lipoprotein subclasses	913	- - -	-5.00[-6.37;-3.63]	8.9×10^{-13}	4807	- - -	-0.92[-1.28;-0.57]	3.7×10^{-7}
Triglycerides to total lipids ratio in very small VLDL	Lipoprotein subclasses	924	- - -	-3.68[-5.07;-2.30]	1.9×10^{-7}	4814	- - -	-1.05[-1.42;-0.68]	3.3×10^{-8}

(b) Metabolites associated with eGFR only in diabetics

Metabolite	Group	T2D				non-diabetics			
		N	signs	β [95%CI]	p-value	N	signs	β [95%CI]	p-value
Esterified cholesterol	Cholesterol	916	+++	4.35 [2.96: 5.74]	9.3×10^{-10}	4799	ooo	0.05 [−0.30: 0.41]	7.6×10^{-1}
Serum total cholesterol	Cholesterol	926	+++	3.68 [2.29: 5.07]	2.0×10^{-7}	4816	ooo	0.03 [−0.32: 0.38]	8.7×10^{-1}
Total cholesterol in LDL	Cholesterol	924	+++	3.84 [2.48: 5.20]	3.4×10^{-8}	4813	ooo	−0.02 [−0.38: 0.33]	9.0×10^{-1}
Cholesterol esters in IDL	Lipoprotein subclasses	926	+++	4.07 [2.71: 5.42]	4.6×10^{-9}	4817	ooo	−0.14 [−0.49: 0.21]	4.3×10^{-1}
Cholesterol esters in large LDL	Lipoprotein subclasses	926	+++	3.63 [2.26: 4.99]	1.8×10^{-7}	4817	ooo	−0.11 [−0.46: 0.24]	5.4×10^{-1}
Cholesterol esters in medium LDL	Lipoprotein subclasses	926	+++	2.98 [1.62: 4.35]	1.9×10^{-5}	4817	ooo	−0.16 [−0.51: 0.20]	3.8×10^{-1}
Cholesterol esters in small LDL	Lipoprotein subclasses	926	+++	3.11 [1.74: 4.48]	8.5×10^{-6}	4817	ooo	−0.09 [−0.45: 0.26]	6.1×10^{-1}
Cholesterol esters to total lipids ratio in large LDL	Lipoprotein subclasses	916	+++	4.18 [2.83: 5.54]	1.5×10^{-9}	4809	ooo	0.26 [−0.09: 0.62]	1.5×10^{-1}
Cholesterol esters to total lipids ratio in medium LDL	Lipoprotein subclasses	907	+++	2.72 [1.33: 4.12]	1.3×10^{-4}	4794	ooo	−0.00 [−0.35: 0.35]	10.0×10^{-1}
Cholesterol esters to total lipids ratio in small LDL	Lipoprotein subclasses	907	+++	2.83 [1.43: 4.22]	6.9×10^{-5}	4795	ooo	−0.02 [−0.37: 0.33]	9.0×10^{-1}
Concentration of IDL particles	Lipoprotein subclasses	926	+++	3.06 [1.70: 4.43]	1.1×10^{-5}	4817	ooo	−0.29 [−0.64: 0.06]	1.0×10^{-1}
Concentration of large LDL particles	Lipoprotein subclasses	926	+++	3.11 [1.74: 4.48]	8.7×10^{-6}	4817	ooo	−0.19 [−0.54: 0.17]	3.0×10^{-1}
Free cholesterol in IDL	Lipoprotein subclasses	926	+++	3.20 [1.82: 4.58]	5.2×10^{-6}	4817	ooo	−0.12 [−0.47: 0.23]	4.9×10^{-1}
Free cholesterol in large LDL	Lipoprotein subclasses	926	+++	3.86 [2.49: 5.22]	3.1×10^{-8}	4817	ooo	−0.01 [−0.36: 0.34]	9.6×10^{-1}
Free cholesterol in medium LDL	Lipoprotein subclasses	926	+++	4.03 [2.67: 5.38]	5.6×10^{-9}	4817	ooo	−0.03 [−0.38: 0.32]	8.8×10^{-1}
Free cholesterol in small LDL	Lipoprotein subclasses	926	+++	3.72 [2.36: 5.08]	7.7×10^{-8}	4817	ooo	0.01 [−0.34: 0.37]	9.4×10^{-1}
Phospholipids in IDL	Lipoprotein subclasses	926	+++	3.27 [1.90: 4.64]	2.8×10^{-6}	4817	ooo	−0.19 [−0.54: 0.16]	2.9×10^{-1}
Phospholipids in large LDL	Lipoprotein subclasses	926	+++	3.84 [2.47: 5.20]	3.4×10^{-8}	4817	ooo	−0.10 [−0.45: 0.25]	5.8×10^{-1}
Phospholipids in medium LDL	Lipoprotein subclasses	926	+++	3.84 [2.49: 5.20]	2.8×10^{-8}	4817	ooo	−0.11 [−0.46: 0.25]	5.6×10^{-1}
Phospholipids in small LDL	Lipoprotein subclasses	926	+++	3.89 [2.53: 5.25]	2.0×10^{-8}	4817	ooo	0.15 [−0.20: 0.51]	4.0×10^{-1}
Total cholesterol in large LDL	Lipoprotein subclasses	926	+++	3.90 [2.53: 5.26]	2.1×10^{-8}	4817	ooo	−0.14 [−0.49: 0.21]	4.4×10^{-1}
Total cholesterol in medium LDL	Lipoprotein subclasses	926	+++	3.76 [2.40: 5.12]	6.1×10^{-8}	4817	ooo	−0.08 [−0.43: 0.27]	6.5×10^{-1}
Total cholesterol in small LDL	Lipoprotein subclasses	926	+++	3.28 [1.91: 4.64]	2.5×10^{-6}	4817	ooo	−0.13 [−0.48: 0.22]	4.7×10^{-1}
Total cholesterol to total lipids ratio in medium LDL	Lipoprotein subclasses	911	+++	3.53 [2.14: 4.91]	6.3×10^{-7}	4795	ooo	0.16 [−0.19: 0.51]	3.6×10^{-1}
Total cholesterol to total lipids ratio in small LDL	Lipoprotein subclasses	910	+++	3.56 [2.17: 4.95]	4.9×10^{-7}	4794	ooo	0.10 [−0.25: 0.45]	5.8×10^{-1}
Total lipids in IDL	Lipoprotein subclasses	926	+++	3.33 [1.96: 4.69]	1.8×10^{-6}	4817	ooo	−0.25 [−0.60: 0.10]	1.7×10^{-1}
Total lipids in large LDL	Lipoprotein subclasses	926	+++	3.53 [2.17: 4.89]	3.8×10^{-7}	4817	ooo	−0.14 [−0.50: 0.21]	4.2×10^{-1}
Total lipids in medium LDL	Lipoprotein subclasses	926	+++	3.26 [1.90: 4.62]	2.8×10^{-6}	4817	ooo	−0.17 [−0.52: 0.18]	3.5×10^{-1}
Total lipids in small LDL	Lipoprotein subclasses	926	+++	3.32 [1.96: 4.69]	1.8×10^{-6}	4817	ooo	−0.07 [−0.42: 0.29]	7.2×10^{-1}

(c) Metabolites associated with eGFR only in non-diabetics

Metabolite	Group	T2D				non-diabetics			
		N	signs	β [95 %CI]	p-value	N	signs	β [95 %CI]	p-value
Ratio of apolipoprotein B to apolipoprotein A-I	Apolipoproteins	925	ooo	0.27 [−1.12: 1.66]	7.0×10^{-1}	4817	---	−1.01 [−1.40: −0.62]	3.0×10^{-7}
Total cholesterol in VLDL	Cholesterol	926	ooo	0.15 [−1.24: 1.53]	8.3×10^{-1}	4817	---	−1.00 [−1.37: −0.63]	1.5×10^{-7}
Lactate	Glycolysis and Gluconeogenesis	914	ooo	0.41 [−1.02: 1.84]	5.7×10^{-1}	4717	---	−0.79 [−1.15: −0.43]	1.6×10^{-5}
Cholesterol esters in chylomicrons and extremely large VLDL	Lipoprotein subclasses	923	ooo	−0.38 [−1.78: 1.02]	5.9×10^{-1}	4805	---	−0.66 [−1.02: −0.29]	3.9×10^{-4}
Cholesterol esters in large VLDL	Lipoprotein subclasses	924	ooo	−0.70 [−2.11: 0.71]	3.3×10^{-1}	4808	---	−0.76 [−1.13: −0.38]	7.1×10^{-5}
Cholesterol esters in medium VLDL	Lipoprotein subclasses	926	ooo	−0.30 [−1.69: 1.08]	6.7×10^{-1}	4816	---	−0.73 [−1.10: −0.37]	9.0×10^{-5}
Cholesterol esters in small VLDL	Lipoprotein subclasses	926	ooo	−0.07 [−1.44: 1.31]	9.2×10^{-1}	4817	---	−1.11 [−1.47: −0.74]	3.2×10^{-9}
Cholesterol esters in very large VLDL	Lipoprotein subclasses	924	ooo	−0.59 [−1.99: 0.82]	4.1×10^{-1}	4804	---	−0.76 [−1.13: −0.39]	5.8×10^{-5}
Concentration of large HDL particles	Lipoprotein subclasses	926	ooo	1.64 [0.14: 3.14]	3.2×10^{-2}	4817	+++	1.06 [0.65: 1.47]	4.8×10^{-7}
Concentration of very small VLDL particles	Lipoprotein subclasses	926	ooo	0.18 [−1.19: 1.55]	8.0×10^{-1}	4817	---	−1.00 [−1.36: −0.64]	3.8×10^{-8}
Free cholesterol in large VLDL	Lipoprotein subclasses	924	ooo	−0.34 [−1.75: 1.08]	6.4×10^{-1}	4806	---	−0.64 [−1.02: −0.27]	7.2×10^{-4}
Free cholesterol in very small VLDL	Lipoprotein subclasses	926	ooo	0.11 [−1.26: 1.47]	8.8×10^{-1}	4817	---	−1.00 [−1.35: −0.65]	2.6×10^{-8}
Free cholesterol to total lipids ratio in large LDL	Lipoprotein subclasses	921	ooo	1.00 [−0.41: 2.40]	1.6×10^{-1}	4803	+++	0.66 [0.28: 1.04]	5.8×10^{-4}
Free cholesterol to total lipids ratio in medium VLDL	Lipoprotein subclasses	916	ooo	−0.51 [−1.92: 0.90]	4.8×10^{-1}	4788	---	−1.06 [−1.41: −0.71]	4.6×10^{-9}
Phospholipids in large VLDL	Lipoprotein subclasses	924	ooo	−0.25 [−1.67: 1.17]	7.3×10^{-1}	4807	---	−0.72 [−1.10: −0.35]	1.7×10^{-4}
Phospholipids in very small VLDL	Lipoprotein subclasses	926	ooo	1.13 [−0.24: 2.51]	1.1×10^{-1}	4817	---	−0.68 [−1.04: −0.33]	1.4×10^{-4}
Phospholipids to total lipids ratio in large HDL	Lipoprotein subclasses	888	ooo	0.37 [−1.10: 1.85]	6.2×10^{-1}	4613	---	−0.81 [−1.22: −0.41]	9.4×10^{-5}
Phospholipids to total lipids ratio in very large HDL	Lipoprotein subclasses	824	ooo	−0.16 [−1.60: 1.27]	8.2×10^{-1}	4676	+++	0.69 [0.32: 1.06]	2.8×10^{-4}
Total cholesterol in chylomicrons and extremely large VLDL	Lipoprotein subclasses	923	ooo	−0.21 [−1.62: 1.19]	7.7×10^{-1}	4804	---	−0.65 [−1.02: −0.28]	5.2×10^{-4}
Total cholesterol in large VLDL	Lipoprotein subclasses	924	ooo	−0.50 [−1.92: 0.91]	4.8×10^{-1}	4809	---	−0.77 [−1.14: −0.39]	6.1×10^{-5}
Total cholesterol in medium VLDL	Lipoprotein subclasses	926	ooo	−0.50 [−1.89: 0.90]	4.8×10^{-1}	4813	---	−0.87 [−1.25: −0.50]	4.4×10^{-6}
Total cholesterol in small VLDL	Lipoprotein subclasses	926	ooo	−0.50 [−1.88: 0.88]	4.8×10^{-1}	4817	---	−1.23 [−1.60: −0.86]	8.8×10^{-11}
Total cholesterol in very large VLDL	Lipoprotein subclasses	924	ooo	−0.28 [−1.69: 1.12]	6.9×10^{-1}	4805	---	−0.71 [−1.08: −0.34]	1.5×10^{-4}
Total lipids in large VLDL	Lipoprotein subclasses	925	ooo	−0.22 [−1.64: 1.19]	7.6×10^{-1}	4813	---	−0.69 [−1.07: −0.31]	3.7×10^{-4}
Total lipids in very small VLDL	Lipoprotein subclasses	926	ooo	0.48 [−0.89: 1.85]	4.9×10^{-1}	4817	---	−0.96 [−1.31: −0.60]	1.4×10^{-7}
Triglycerides in large LDL	Lipoprotein subclasses	926	ooo	−0.67 [−2.05: 0.72]	3.5×10^{-1}	4816	---	−0.85 [−1.20: −0.50]	2.3×10^{-6}
Triglycerides in large VLDL	Lipoprotein subclasses	924	ooo	−0.14 [−1.56: 1.28]	8.5×10^{-1}	4811	---	−0.65 [−1.03: −0.27]	7.6×10^{-4}
Triglycerides in medium LDL	Lipoprotein subclasses	926	ooo	−0.54 [−1.92: 0.85]	4.5×10^{-1}	4815	---	−0.78 [−1.13: −0.42]	1.6×10^{-5}
Triglycerides in small LDL	Lipoprotein subclasses	926	ooo	−0.47 [−1.86: 0.92]	5.1×10^{-1}	4815	---	−0.68 [−1.04: −0.32]	1.9×10^{-4}

6.3 Discussion

In the largest study of its kind, including 926 T2D diabetics and 4838 non-diabetics from four independent European cohorts, I identified 142 metabolites significantly associated with renal function: 59 in diabetics, 109 in non-diabetics, with an overlap of 26 metabolites. While associations were largely concordant ($R^2 = 0.60$, Figure 6.2), there were some notable exceptions. For instance, the amino acids glycine and phenylalanine as well as the energy metabolites citrate and glycerol were all negatively associated with eGFR in diabetics and non-diabetics, while lipidomics profiles were rather different between the groups.

6.3.1 Renal production of amino acids

Phenylalanine is an essential, aromatic amino acid, which serves as precursor for tyrosine in the liver and kidneys (Møller et al., 2000) as well as dopamine-related neurotransmitters (Fernstrom et al., 2007). Phenylalanine has been previously associated with insulin resistance and increased risk for diabetes (Rhee et al., 2011; Wang et al., 2011b; Floegel et al., 2013). Moreover, reduced rates of the conversion of phenylalanine to tyrosine were found in renal disease (Young et al., 1973; Boirie et al., 2004), leading to decreased circulating levels of tyrosine and increased levels of phenylalanine in CKD patients. While previous studies found the negative association of eGFR with tyrosine stronger than its positive associations with phenylalanine (Kopple, 2007), I found tyrosine levels significantly decreased only in diabetic patients with advanced renal disease. In contrast, increased concentration of phenylalanine was observed in both diabetics and non-diabetics, regardless of disease progression in this study. This increase was specific to renal damage, and showed little association with retinopathy.

Interestingly, phenylalanine has been shown to predict cardiovascular events in a multi-centre European study (Wurtz et al., 2015), however, the causal relationship remains elusive. As phenylalanine was not significantly associated with longitudinal change of eGFR, it might be either a consequence of renal damage or a more general marker for biological ageing not specific to renal function.

Similarly, glycine is converted to serine in the kidneys (Pitts et al., 1970; Tizianello et al., 1980). Impairment of renal function leads to decreased conversion of glycine and thus, accumulation of glycine in blood, which I observed consistently in both diabetic and non-diabetic cohorts. In contrast to phenylalanine, glycine also significantly correlated with albuminuria and diabetic retinopathy, but did not predict longitudinal change of renal function.

6.3.2 Energy-related metabolites

Alanine is a major precursor of hepatic and renal gluconeogenesis and glycolysis via deamination to pyruvate, catalysed by ALT in the liver as part of the Cahill cycle. Similarly, the Cori cycle transports lactate produced in the muscle to the liver, where it is also converted to glucose via pyruvate (Berg et al., 2002). Together with glycerol, which also was negatively associated with eGFR only in non-diabetics, and glutamine they compose the substrates of 90 % of the human gluconeogenesis. This indicates accumulation of metabolic end products of

energy metabolism in muscle and decreased gluconeogenesis with advancing kidney disease on non-diabetics. While these processes were traditionally thought to take place mainly in the liver, the kidneys play a major role in glucose metabolism, both releasing and consuming glucose (Schoolwerth et al., 1988). Previous studies demonstrated that metabolic acidosis induced by CKD leads to increased abundance of circulating alanine, glutamine, and glutamate, possibly due to abnormal nitrogen utilization and accelerated net protein breakdown. Moreover, insensitivity to epinephrine observed in uraemia affects the increases of alanine and glutamine (May et al., 1987; Garber, 1978).

In contrast to the observed associations in non-diabetic individuals, glucose metabolism is already heavily disturbed in diabetics with an increased rate of gluconeogenesis (Magnusson et al., 1992). Consequently, the decline of renal function does not have a major impact on gluconeogenesis in diabetics.

6.3.3 Lipoprotein composition

One of the largest difference in metabolic associations with renal function between diabetics and non-diabetics were the negative associations of several VLDL subclasses with eGFR observed only in non-diabetics and the positive associations of IDL and LDL subclasses with eGFR observed only in diabetics.

Dyslipidaemia is a commonly observed in renal disease (Tsimihodimos et al., 2008) and particularly LDL cholesterol is a very well-established risk factor for complications of CKD, including cardiovascular diseases (Rahilly-Tierney et al., 2009) and increased mortality (Lewington et al., 2007). While this is in line with my findings in the non-diabetic cohorts, it is opposite of what I observed in the diabetics. Also, absolute levels of total cholesterol and LDL cholesterol levels were higher in the general population than in the diabetic GenoDiabMar cohort (Figure C.1a), despite the much higher BMI and lower renal function (Table 6.1). This is very likely an effect of lipid-lowering medication, in particular statins, taken by diabetics. The unexpected positive associations of LDL subclasses might be due to stronger medication in patients with worse general health, leading to lower LDL levels associated with lower eGFR.

Interestingly, in contrast to LDL, triglyceride ratios were negatively associated with renal function consistently between diabetics and non-diabetics. Also, triglyceride to total lipid ratios were higher in the GenoDiabMar cohort than in the non-diabetic TwinsUK population (Figure C.1b). While triglyceride levels also respond to statin treatment, this is a secondary effect that follows the decrease of LDL and total cholesterol (Stein et al., 1998). Accordingly, the triglyceride to total lipid ratio is not affected by statins (Würtz et al., 2016). This indicates that reduction of absolute levels of triglycerides might not be as relevant as reducing their relative abundance compared to the overall lipid composition.

6.3.4 Phospholipids in HDL

The ratio of phospholipids to total lipids very large HDL was associated with eGFR, longitudinal change of eGFR, and predicted future eGFR in non-diabetics, however, did not show any association with eGFR in diabetics.

Phospholipids in HDL were found to enhance the cholesterol efflux capacity of HDL (Fournier et al., 1997; Agarwala et al., 2015). Cholesterol efflux is part of the reverse transport of cholesterol to the liver, which is thought to be the cause of the beneficial effect of HDL cholesterol (Tall, 1998). While this appears to be the case in the non-diabetic cohorts, cholesterol efflux is impaired in diabetics (Apro et al., 2016; Zhou et al., 2008). As phospholipids in HDL are not affected by statin treatment (Würtz et al., 2016), this is likely a difference of HDL metabolism between diabetics and non-diabetics, rather than a treatment effect. My results confirm previous findings on the regulation of cholesterol efflux and their differences in diabetics and non-diabetics.

6.4 Conclusion

I analysed associations of 227 metabolic traits with renal function in 5764 individuals from four European cohorts, including 926 type 2 diabetics. I identified 109 metabolites correlated with eGFR in non-diabetics and 59 metabolites in diabetics, 26 of which overlapped between the groups.

I found renal conversion of the amino acids phenylalanine and glycine equally impaired in diabetics and non-diabetics. Also, higher triglyceride content was consistently associated with worse and higher HDL levels with better renal function.

In contrast, the amount of phospholipids in HDL was renoprotective in non-diabetics but not in diabetics, potentially due to their effect on cholesterol efflux. More importantly, decline of renal function heavily affected energy metabolism, particularly gluconeogenesis, in non-diabetics, while it had no effect on the already defective glucose metabolism of diabetics.

I want to note some limitations of this study. The GenoDiabMar cohort is a cohort of diabetic patients, recruited in a hospital, and thus quite different from the other cohorts, which have been recruited from the general population. Most importantly, renal function is on average much worse than in the general population cohorts and all the participants receive various treatments. Thus, some of the observed differences between diabetic and non-diabetic cohorts might not be due to the diabetes status alone but other factors, such as medication. While trying to counteract this effect by replicating the results in two independent cohorts of diabetics, power of the KORA and TwinsUK cohorts was rather limited. Also, treatment of diabetics particularly with statins has a big impact on lipid profiles, which might explain the unexpected associations of LDL. Hence, future analyses should consider the different treatments when analysing metabolic associations with renal function in disease cohorts. Finally, while there were many metabolites

associated cross-sectionally with eGFR very few showed a predictive potential, indicating that metabolic changes might be consequences rather than causes of kidney disease. However, this might be due to the limited precision of creatinine.

In summary, I found widespread metabolic changes associated with decline of renal function. While changes of lipid composition, particularly triglycerides and HDL cholesterol, were largely similar between diabetics and non-diabetics, changes of energy metabolism were markedly different. Most of the observed changes are likely consequences of CKD, which might cause complications of renal disease, such as cardiovascular events.

CHAPTER 7

The faecal metabolome as a functional readout of the gut microbiome

In this chapter, I analysed the faecal metabolome, assessing its potential as novel profiling method for the gut microbiome and a potential marker of biological ageing. I calculated its heritability, conducted a GWAS, and finally associated it with the gut microbial community to quantify host and gut microbial influences. I then used the faecal metabolome to investigate associations of microbial metabolism with age, obesity, and renal function.

There is growing awareness that the gut microbiome has a beneficial role in maintaining homeostasis of host metabolism (O'Hara et al., 2006) and that disruption of this intricate system is implicated in human health. The overall microbial composition as well as individual microbial species have been associated with a range of diseases such as obesity (Ley et al., 2006; Turnbaugh et al., 2009; Beaumont et al., 2016), insulin resistance (Pedersen et al., 2016), and inflammatory bowel disease (IBD) (Kostic et al., 2014). While the microbiome was shown to be relatively stable during most of adulthood (Turnbaugh et al., 2009; Costello et al., 2009), it strongly associates with frailty (Jackson et al., 2016a) and changes dramatically in late life (Claesson et al., 2012), suggesting an involvement in the ageing process, possible due to its influence on inflammatory processes (Ouwehand et al., 2008).

Metabolomics and the gut microbiome are strongly related, with microbes producing many of the body's chemicals, hormones, and vitamins (Clarke et al., 2014). Research has shown that the gut microbiome has an effect on circulating levels of several metabolites. For example, increased microbial production of branched-chain amino acids (BCAAs) together with decreased microbial uptake was shown to increase serum levels of BCAAs, potentially causing insulin resistance (Pedersen et al., 2016). There is also some evidence that microbial metabolites, such as p-cresol sulfate and indoxyl sulfate, are involved in the decline of renal function (Ramezani et al., 2014; Barrios et al., 2015).

However, to date most microbiome studies rely on sequencing rather than metabolomics data. Despite the advances of next generation sequencing platforms, which facilitate profiling of complex microbial communities using 16S sequencing or metagenomics (Jovel, 2016), annotation is sparse at the species level and virtually absent at the strain level. Moreover,

the microbiome only codes microbial possibilities rather than their actual activity and can, for instance, not differentiate between alive and dead microbes (Cangelosi et al., 2014), nor determine the transcriptional activity of the genes within each bacterial genome (Frias-Lopez et al., 2008).

Faecal metabolomics, on the other hand, reports specifically on the metabolic interplay between the host, diet, and the gut microbiota (Marcobal et al., 2013). A comparison of conventional and germ-free mice demonstrated the strong impact of the microbiome on intestinal metabolite profiles (Matsumoto et al., 2012). Later studies confirmed strong dependencies of the microbiome with gut metabolite concentrations in humans (McHardy et al., 2013). Thus, faecal metabolomics might provide a functional readout of the gut microbiome that complements sequencing-based methods. Claesson et al. (2012), for instance, identified marked changes in the gut microbiome of long-term residential care patients and then used metabolomics measurements from a subset of the patients to identify significant changes of faecal lipid metabolism. While the faecal metabolome has been shown to add valuable information to sequencing data in small case-control studies (Choo et al., 2015; De Preter et al., 2015; Antharam et al., 2016) and model organisms (Ng Hublin et al., 2013), to date there is no systematic or large-scale analysis of the faecal metabolome in the general population.

The aim of this study is to provide, for the first time, a comprehensive description and assessment of the potential utility of the faecal metabolome in a large population-based setting. To this end, I analysed faecal metabolomics measurements from 786 individuals and (i) their associations with ageing, obesity, and renal function (ii) host genetic influences and (iii) multivariate dependencies with the gut microbiome.

7.1 Methods

7.1.1 Study population

Study participants were 786 predominantly female (93%) twins from the TwinsUK cohort with faecal metabolomic profiles available. 739 of them additionally had whole genome sequencing data available and 644 of the faecal samples were also profiled using 16S sequencing. Measurements of visceral fat were available for 652 and measurements of renal function for 613 individuals (Figure 7.1).

7.1.2 Data collection

Faecal samples were collected from study participants and stored at -80°C . 16S sequencing of faecal samples was conducted by Cornell University as described in section 3.3.3. Metabolomic profiling of faecal samples was conducted by Metabolon, using the version V4 of their untargeted platform as described in section 3.3.1.

Table 7.1 Population characteristics.

Faecal samples from 786 individuals of the TwinsUK cohort were profiled by Metabolon Inc. Whole genome sequencing (WGS), 16S sequencing and measurements of visceral fat and renal function were available for subsets of the study population.

TwinsUK	
Sample size, <i>n</i>	786
MZ:DZ:singletons	296:310:180
Age, years	65.2 ± 7.6
Female, <i>n</i> (%)	734 (95.4%)
BMI, kg/m ²	26.1 ± 4.7

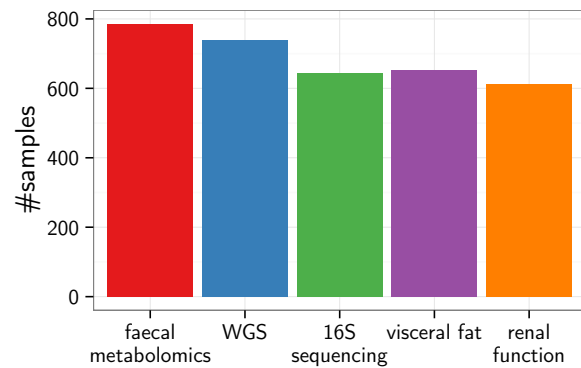
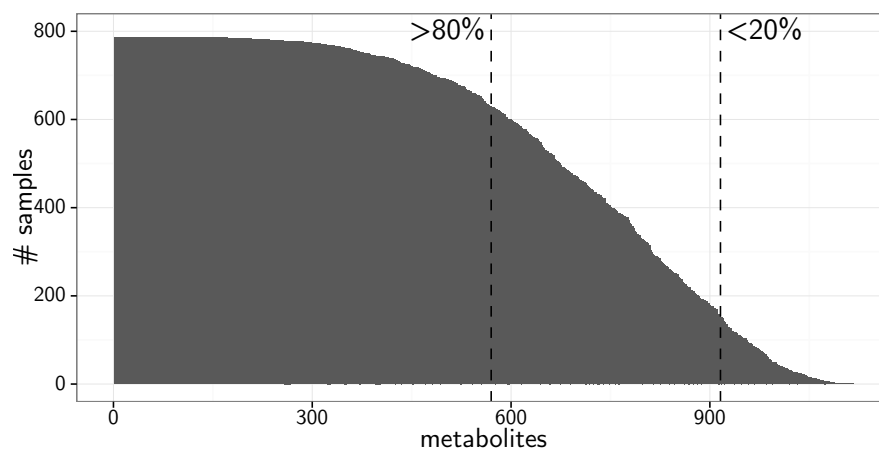
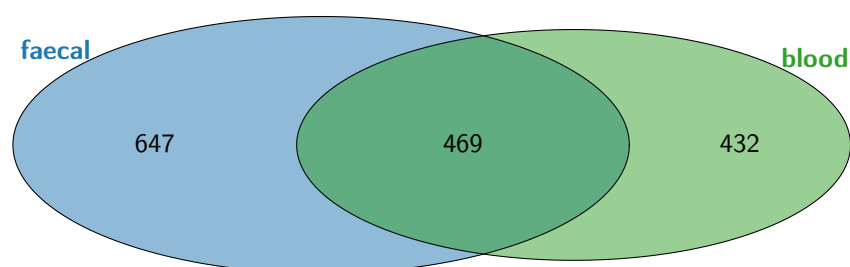
Figure 7.1 Study population.

Figure 7.2 The frequency of faecal metabolites in TwinsUK. 1116 metabolites were detected in 786 faecal samples. (a) 570 of those were detected in at least 80% of all samples and 345 were detected in less than 80% but more than 20% of all samples. The first were analysed continuously, while I dichotomized the latter in present/absent. 210 metabolites that were present in less than 20% of the samples were excluded from further analysis. (b) 469 metabolites were observed in both faecal and blood samples of the same individuals, while 647 metabolites were only measured in faeces. 499 of these 647 faeces-specific metabolites were observed in at least 20% of the samples.

(a) Faecal Metabolites**(b) Overlap of faecal and blood metabolites**

A total of 1116 different metabolites were measured in the 786 faecal samples, of which 210 were observed in less than 20 % of the samples and were therefore excluded from further analyses, due to the limited power in subsequent analyses (Figure 7.2). 345 metabolites were observed in more than 20 % but less than 80 % of the samples and were analysed qualitatively as dichotomous trait (observed in a sample vs. not observed). The remaining 570 metabolites, which were observed in at least 80 % of all samples, were scaled by run-day medians, log-transformed and scaled to uniform mean 0 and SD 1 and analysed quantitatively, as described before (see section 3.3.1).

To analyse the effect of sample storage time on faecal metabolite profiles, I regressed each metabolite against the time (i) in the fridge before samples arrived in our biobank, and (ii) in the freezer before being further analysed. After correction for multiple testing ($FDR < 5\%$) I found significant storage effects on 7 metabolites (Figure D.1). Consequently, I corrected all further analyses for both storage time in the fridge and freezer, to avoid spurious results.

7.1.3 Statistical analysis

7.1.3.1 Associations with phenotypes

To investigate the dependency of the faecal metabolome on chronological age, I first regressed each metabolite against age, correcting for sex, storage times, and family structure as random intercept (section 3.4.2). Then, I calculated linear and logistic regression models, respectively, to assess the relationship of the faecal metabolome with obesity, measured as BMI (section 3.2.1) and visceral fat mass (section 3.2.2), adjusting for age, sex, storage time, and family as random intercept. Lastly, I regressed faecal metabolites against the eGFR, calculated from serum creatinine (section 3.2.3), to assess their association with renal function. Pathway enrichment tests were performed as described before using *piano* (see section 3.4.3).

The faecal metabolome is expected to be influenced by dietary changes (Ríos-Covián et al., 2016; Russell et al., 2011), which also influence the microbiome composition, as well as host phenotypes. Thus, diet is a potential cause of the observed associations between faecal metabolites and microbes, as well as faecal metabolites and host phenotypes. There is also a potential interaction effect between the abundance of nutritional compounds and gut microbes on the presence of faecal metabolites. These complex dependencies between nutrition, faecal metabolites, and microbes were not subject of this analysis and should be addressed in future studies with comprehensive data on nutrition available, ideally 24-hour dietary recalls.

7.1.3.2 Host genetic influences

I used ACE models to estimate the heritability and common environment components of 915 faecal metabolites. Liability threshold models were used for dichotomous metabolic traits (section 3.4.1).

Genetic variation was measured using whole genome sequencing (see section 3.3.4). I excluded 127,778 SNPs with Hardy-Weinberg $p < 1.0 \times 10^{-6}$, calculated from 420 unrelated individuals, leaving 8,222,692 biallelic SNPs with minor allele frequency higher than 1 % for further analysis (see section 3.4.2.3). To test for associations of the 428 heritable faecal metabolites with SNPs, I fitted linear mixed models correcting for age, sex, storage time, and the sample kinship using GEMMA (Zhou et al., 2012). I used Bonferroni-correction to account for multiple testing, adjusting for 428 tested metabolites ($p < 1.2 \times 10^{-10} = 5.0 \times 10^{-8} / 428$).

Additionally, I tested for genetic associations with 31,226 pairwise metabolite ratios of faecal metabolites with known chemical identity and a heritable variance component. I used the *p*-gain statistic to assess independence of the individual metabolites (Petersen et al., 2012). The *p*-gain is defined as the minimal *p*-value of the associations of either of the individual metabolites divided by the *p*-value of the metabolite ratio. A high *p*-gain statistic indicates that the ratio carries additional information compared to individual metabolites. I considered metabolite ratios with $p < 1.6 \times 10^{-12}$ ($= 5.0 \times 10^{-8} / 31,226$ metabolite ratios) and $p\text{-gain} > 3.1 \times 10^5$ ($= 10 \times 31,226$ metabolite ratios) significant.

7.1.3.3 Associations of the faecal metabolome with the gut microbiome

To assess associations of the faecal metabolome with the gut microbiome composition, I first regressed metabolite concentrations against the Shannon alpha diversity (section 3.3.3), adjusting for age, sex, BMI, storage time and family structure using 644 individuals with both faecal metabolomics and 16S sequencing data available.

In the second step, I estimated the proportions of variance explained by the microbiome for each metabolite. To this end, I scaled UniFrac beta-diversity distances (see section 3.3.3) to [0; 1] and converted them to a similarity matrix by subtracting each entry from 1. I then calculated restricted maximum likelihood (REML) models, regressing each metabolite level against the microbial similarity, adjusting for age, sex, BMI, storage times and technical covariates (sequencing run, sequencing depth, individual who extracted the DNA, individual who loaded the DNA and sample collection method) using the R package *regress*. The proportion of variance explained by microbial similarity and its standard error were calculated from the variance components using the R package *gap* and *p*-values were calculated from the ratio of the estimate over its standard error. This technique is commonly used to estimate the heritability from genetic kinship matrices (Yang et al., 2011; Speed et al., 2012), however, is much less common in microbiome research.

Next, I aimed to identify individual microbial OTUs and higher taxonomical units that are associated with faecal metabolite levels. To this end, I regressed 581 inverse normalised OTUs against all 915 metabolites, adjusting for age, sex, BMI, sample storage times, family structure and the alpha diversity. I further calculated associations of higher taxonomical units, ranging from genus to phylum level. Benjamini-Hochberg correction was applied to correct each of these analyses for multiple testing.

Lastly, to assess multivariate dependencies between the faecal metabolome and the microbiome, I inferred a GGM combining 423 metabolites with known chemical identity that were observed in at least 80 % of the samples with 241 OTUs that were assigned complete taxonomy at least to the genus level (see section 3.4.7).

7.2 Results

The faecal samples of 786 predominantly female participants of the TwinsUK cohort, aged 65.2 (± 7.6), with an average BMI of 26.1 (± 4.7) (Table 7.1) were analysed by the metabolomics provider Metabolon Inc. using an untargeted MS platform. A total of 1116 metabolites, 849 of them with known chemical identity were measured.

Among the detected metabolites, 570 were found in at least 80 % of the samples, while 345 were detected in at least 20 % but in less than 80 % of all samples (Figure 7.2). The latter were analysed as dichotomous traits (present/absent in a sample) and metabolites measured in less than 20 % of the samples were discarded from further analysis. 647 faecal metabolites (499 of them present in at least 20 % of samples) were not detected in blood samples of the same individuals that were profiled on the same platform (Figure 7.2b).

7.2.1 Faecal metabolite associations with phenotypes

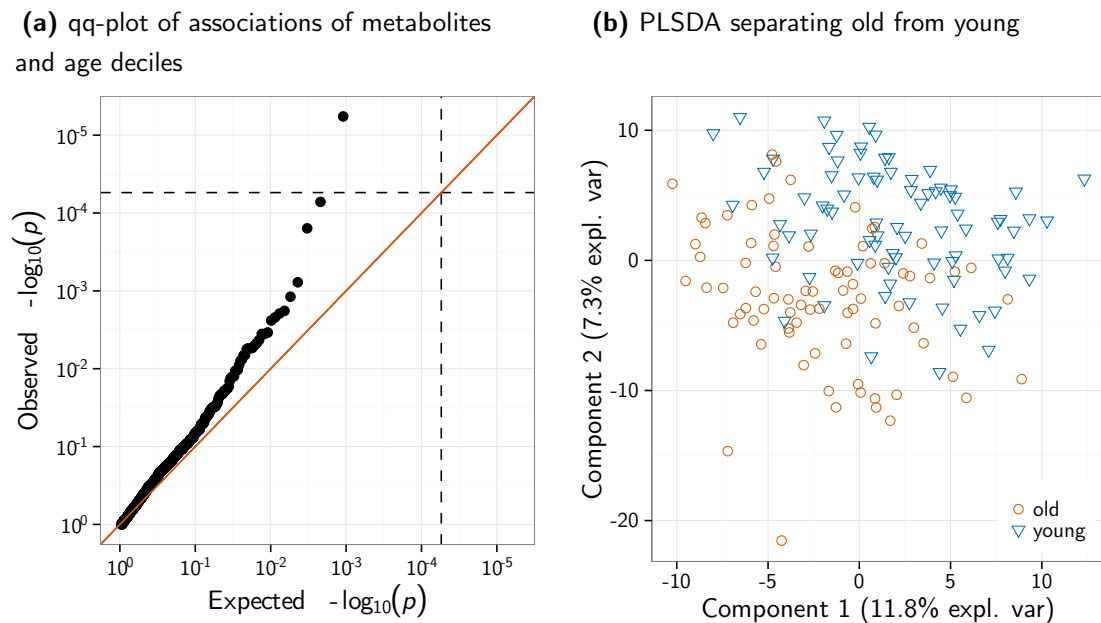
I tested 915 faecal metabolites for association with adult age and did not find significant associations after correcting for multiple testing (FDR adjusting for 3660 tests = 915 metabolites \times 4 phenotypes). While faecal metabolites were not linearly associated with adult age, I found differences between the oldest decile (>75 years) and the youngest decile (<56 years) of the study population (Figure 7.3). A multivariate partial least squares discriminant analysis (PLSDA) incorporating all 570 common metabolites could distinguish the two extremes (AUC = 71 %, $p = 6.8 \times 10^{-6}$, Figure 7.3b). This separation was partly driven by phytanate ($\beta[95\%CI] = -0.68 [-0.97 : -0.39]$, $p = 5.7 \times 10^{-6}$), the only metabolite that I found significantly different between the age groups using regression models (Figure 7.3a).

BMI was found to be associated with eight metabolites at a FDR of 5 %: five faecal lipids, including arachidonate ($\beta[95\%CI] = 0.13 [0.07 : 0.19]$, $p = 1.1 \times 10^{-5}$), the haemoglobin metabolite bilirubin ($\beta[95\%CI] = 0.13 [0.06 : 0.19]$, $p = 8.9 \times 10^{-5}$) and two unknown metabolites (Table D.1).

Moreover, I found one faecal amino acid significantly associated with eGFR after corrections for multiple testing: 1-methylguanidine ($\beta[95\%CI] = -2.12 [-3.06 : -1.19]$, $p = 8.6 \times 10^{-6}$).

Lastly, I looked for associations with visceral fat mass, a measure of abdominal obesity, and found a total of 102 statistically significant associations (FDR $< 5\%$), 13 of them significant also after adjusting for stringent Bonferroni correction (again adjusting for 915×4 tests)

Figure 7.3 Faecal metabolites differ between the young and the elderly. While I found the faecal metabolome stable during adulthood, the oldest decile (>75 years) and the youngest decile (<56 years) of the study population were significantly different. (a) I first investigated the age effect for all metabolites individually and found one metabolite, phytanate, significantly different between the age groups. (b) I then fitted a multivariate PLSDA model to distinguish the older from the younger group. The AUC was estimated at 0.71 ($p = 6.8 \times 10^{-6}$) in a 10-fold cross-validation setting.



(Table D.1). Visceral fat associated metabolites were significantly enriched for N-acetyl-amino acids (43 metabolites, enrichment p -value $< 2.0 \times 10^{-4}$). Moreover, I found 14 fatty acids, including arachidonate ($\beta[95\%CI] = 5.07 [2.55:7.59]$, $p = 8.2 \times 10^{-5}$, 8 nucleotides, 6 sugars and 6 vitamins associated with visceral fat. Notably, faecal concentrations of four B-vitamins – nicotinate ($\beta[95\%CI] = 7.38 [4.91 : 9.86]$, $p = 5.1 \times 10^{-9}$ and pantothenate ($\beta[95\%CI] = 6.12 [3.60:8.64]$, $p = 1.9 \times 10^{-6}$), riboflavin ($\beta[95\%CI] = 4.61 [2.16:7.07]$, $p = 2.3 \times 10^{-4}$) and pyridoxamine ($\beta[95\%CI] = 3.84 [1.40 : 6.28]$, $p = 2.1 \times 10^{-3}$ – were positively associated with visceral fat.

7.2.2 Host genetic influence on the faecal metabolome

Heritability estimates were computed using SEM (see section 3.4.1). For 428 metabolites, the best fitting model contained a heritable variance component (A), which accounted on average for 17.9% ($\pm 9.7\%$) of the metabolite variation. I found the abundance of long-chain fatty acids, such as 1-palmitoyl-2-arachidonoyl-GPC ($H^2 = 60.7\% [43.4 : 78.0]$) and stearylcarntine ($H^2 = 54.3\% [36.4 : 72.3]$), amongst the most heritable metabolites. For 279 metabolites, including the coffee-metabolite 5-acetyl-amino-6-amino-3-methyluracil ($C = 30.3\% [20.0 : 40.6]$), the best fitting model was the CE model, with an average common environment component (C) of 14.8% ($\pm 8.1\%$). For the remaining 208 metabolites, the best fitting model was the E model, indicating

Figure 7.4 Variance components of the faecal metabolome. I estimated heritability (red) and the effect of common environment (blue) using ACE models from 148 pairs of MZ twins and 155 pairs of DZ twins. Additionally, I estimated the proportion of variance explained by the gut microbiome (green) using mixed models incorporating the microbial beta-diversity for 644 individuals with 16S sequencing data available. The heat-map panels show associations of faecal metabolites with microbial alpha diversity ($n = 644$, red), visceral fat mass ($n = 647$, blue) and BMI ($n = 786$, green), where darker colours indicate stronger associations and grey indicates non-significant associations (FDR corrected). Results for metabolites observed in less than 80 % of the samples are presented in Figure D.2.

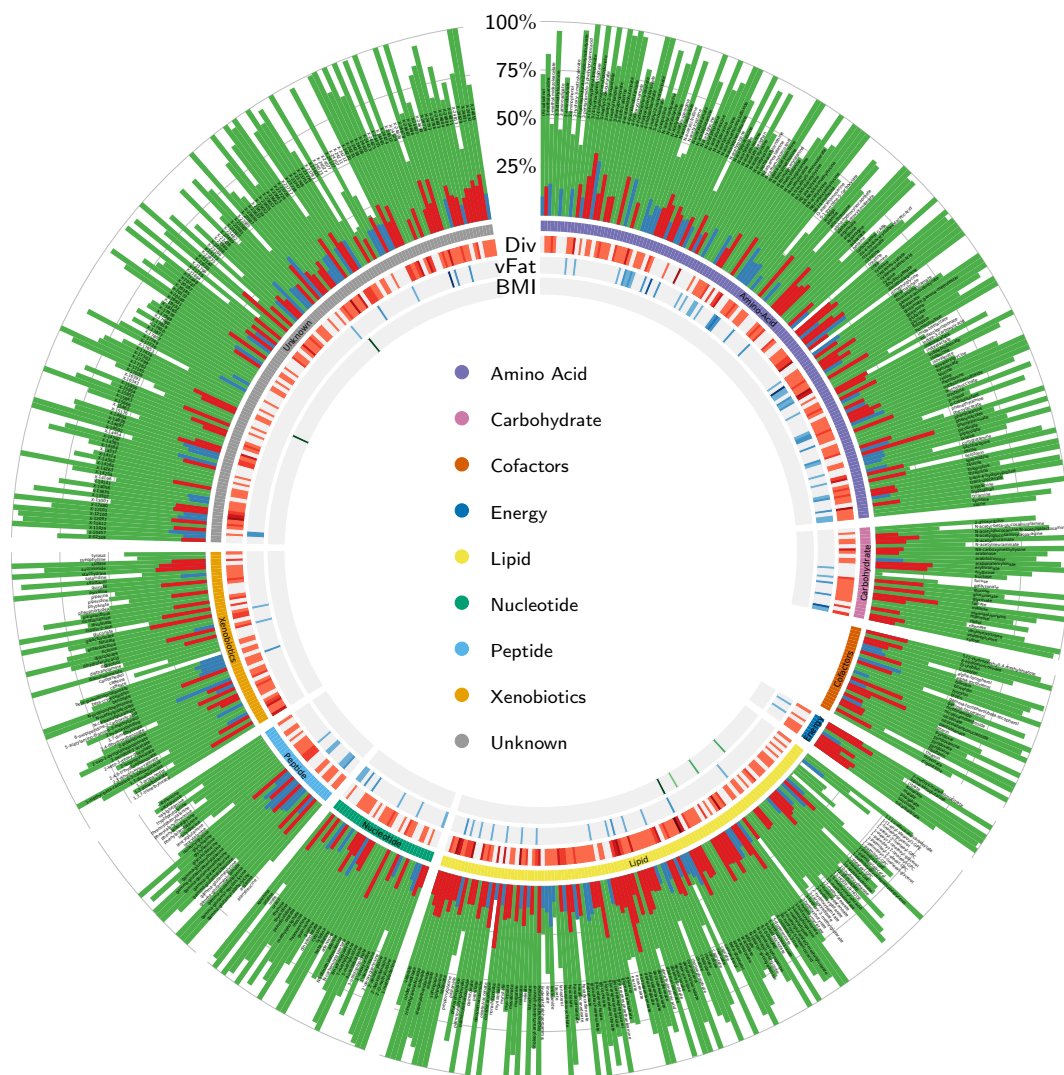
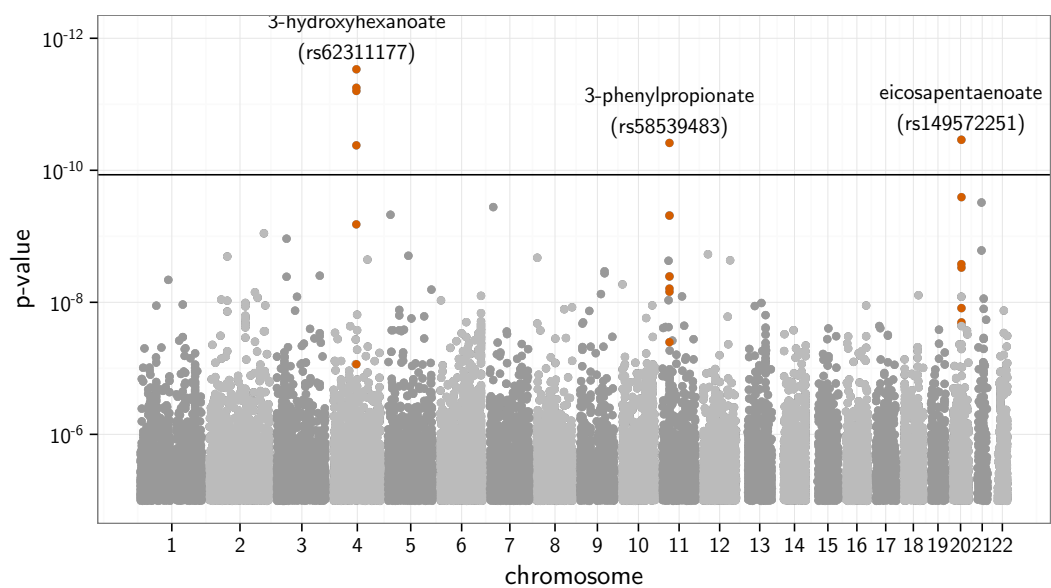
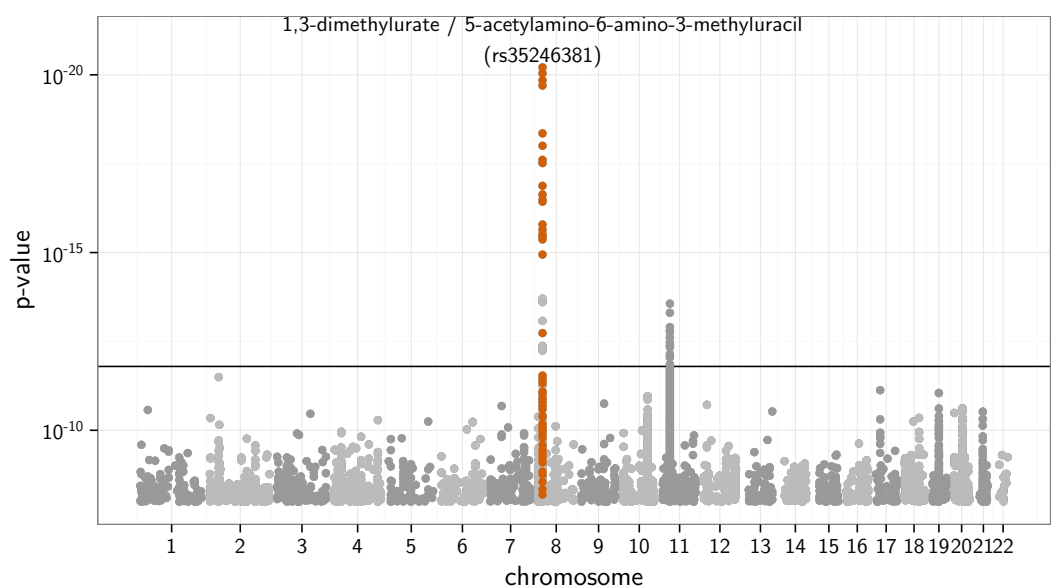


Figure 7.5 Manhattan plot of faecal metabolites. (a) I found three metabolites significantly associated with genetic loci. The horizontal line indicates the Bonferroni cut-off of 1.2×10^{-10} . Significant loci are coloured in orange. (b) Additionally, two metabolite ratios passed Bonferroni correction ($p < 1.6 \times 10^{-12}$): 1,3-dimethylurate / 5-acetylamino-6-amino-3-methyluracil ($p = 6.2 \times 10^{-21}$) and 3-phenylpropionate (hydrocinnamate) / benzoate ($p = 1.5 \times 10^{-15}$). However, only the first one passed filtering by p -gain ($p\text{-gain} > 8.9 \times 10^5$) and thus was considerably stronger than the association of each individual metabolite.

(a) Individual metabolites



(b) Metabolite ratios



that the entire variation of the metabolite is due to individual differences such as the microbiome or diet (Table D.1, Figures 7.4 and D.2). I found a significantly stronger environmental effect on lipids than other metabolites (enrichment $p < 2.0 \times 10^{-4}$).

To investigate the influence of host genetics on the faecal metabolome, I conducted GWASs for the 428 metabolites with a heritable variance component (Figure 7.5a). Three metabolites were significantly associated with genetic loci after correcting for multiple testing ($p < 1.2 \times 10^{-10} = 5.0 \times 10^{-8} / 428$ tested metabolites) (Table 7.2). These include the amino acid 3-phenylpropionate (rs58539483, $p = 2.3 \times 10^{-11}$, Figure D.4a) and the two lipids eicosapentaenoate (rs149572251, $p = 4.7 \times 10^{-11}$, Figure D.4b) and 3-hydroxyhexanoate (rs62311177, $p = 8.6 \times 10^{-12}$, Figure D.4c).

I also tested for genetic associations of metabolite ratios (Figure 7.5b), which were shown to be often better proxies for chemical reactions than single metabolites (Petersen et al., 2012). After correcting for 31,226 tested ratios ($p < 1.6 \times 10^{-12}$), I found the ratio of 5-acetylamino-6-amino-3-methyluracil over 1,3-dimethylurate associated with a locus on chromosome 8 ($p = 5.2 \times 10^{-20}$, $p\text{-gain} = 1.8 \times 10^9$, Figure D.4d). Regional association plots and qq-plots of these associations are shown in Figures D.3 and D.5, respectively.

7.2.3 Microbial association with the faecal metabolome

As the faecal metabolome appeared to be moderately influenced by host genetics, I hypothesised that it reflects to a large extent metabolic processes of the gut microbiome. I regressed metabolite levels against microbial diversity, quantified by the Shannon index, and found that more than the half of the metabolites (575 metabolites) across all pathways were significantly association with diversity (FDR < 5 %), 347 of them passing a conservative Bonferroni correction.

I then estimated the proportion of variance explained by microbiome composition for each metabolite, using the UniFrac beta-diversity metric, a measure of overall phylogenetic dissimilarity between individuals' microbiota (Lozupone et al., 2005). I found that gut microbial composition explained a significant proportion of the observed variance of 710 metabolites, on average 67.7 % (± 18.8 %), ranging from 22.1 % for 1-linolenoylglycerol up to 100 % for several N-acetyl-amino acids (Table D.1). Amongst others, the microbiome explained a significant

Table 7.2 Genetic loci associated with faecal metabolites.

Metabolite	h^2	SNP	Position	Gene	beta	p	p-gain
1,3-dimethylurate / 5-acetylamino- 6-amino-3- methyluracil	39.8	rs35246381	8:18415025	NAT2	-0.17[-0.21:-0.14]	6.2×10^{-21}	7.5×10^9
3-hydroxyhexanoate	22.3	rs62311177	4:92962004	GRID2	0.41[0.30: 0.52]	3.0×10^{-12}	
3-phenylpropionate	24.3	rs58539483	11:28830501	RP11- 115J23.1	-1.31[-1.68:-0.94]	3.9×10^{-11}	
eicosapentaenoate	16.0	rs149572251	20:34322936	ITCH	1.45[1.04: 1.86]	3.5×10^{-11}	

proportion of the variance of all the 8 BMI-related, and 101 of the visceral fat-related metabolites. Xenobiotics, including the B-vitamins nicotinate and pantothenate, showed the strongest associations with microbial composition (enrichment $p < 1.0 \times 10^{-4}$).

To explore the associations of the faecal metabolome with gut microbes at a finer taxonomic resolution, I regressed each metabolite against the 581 most abundant OTUs, adjusting for potential confounding factors including Shannon diversity. I found 42,645 significant associations of 907 different metabolites with 579 different OTUs after adjusting for multiple testing (FDR, correcting for 531,615 tests = 915 metabolites \times 581 OTUs). Moreover, I calculated associations of faecal metabolites with collapsed taxonomical levels, ranging from genus to phylum level (Figure 7.6). 264 metabolites only associated with microbes at the OTU level, while the remaining metabolites were also associated with broader taxonomic groupings.

Lastly, to investigate the connectivity of the faecal metabolome with microbes, I inferred a GGM combining 423 common metabolites with known chemical identity and 241 OTUs with complete taxonomy assignment to at least genus level. The resulting model consisted of 2460 independent associations, 1000 of them amongst metabolites, 921 amongst microbes, and 539 connecting metabolites and microbes. Even though the network is much sparser than a correlation network (1.1 % of all edges in the network compared to 37.9 % in correlation network), all but 13 variables form one connected component. I detected 19 clusters in the largest component, 10 of which contained both microbes and faecal metabolites and 9 consisted of metabolites only. Xenobiotics had higher node degrees (enrichment $p < 3.0 \times 10^{-4}$) and were more densely connected with OTUs ($p < 2.4 \times 10^{-3}$). However, the most connected metabolite was carboxyethyl-GABA, a neurotransmitter that is produced by microbes (Cryan et al., 2012). The most densely connected OTUs in the network were assigned to the *Clostridiales* order.

7.3 Discussion

This is the first comprehensive description of the faecal metabolome. I analysed over 1000 metabolites from the faecal samples of 786 individuals from the TwinsUK cohort. I show that the faecal metabolome is stable during adulthood, influenced by host genetics, and strongly associated with the gut microbial community. Indeed, it provides a functional readout of microbial metabolism and constitutes a unique opportunity to study mechanisms of microbial interaction with host phenotypes. As an example, I found strong associations of specific faecal metabolites with abdominal obesity that provide functional information backing up previous strong but crude associations of the microbiome on obesity (Turnbaugh et al., 2009).

7.3.1 Faecal metabolites are stable during ageing

While the faecal metabolome was stable during adulthood, I found the oldest decile of the study population significantly different from younger individuals, suggesting that the faecal metabolome changes late in life. This is in line with previous findings on the effects of age on the gut microbiome (O'Toole et al., 2010; Claesson et al., 2011; Yatsunenko et al., 2012).

Figure 7.6 Associations of faecal metabolites with gut microbes. Associations of faecal metabolites and the gut microbiome are represented as Manhattan plot, where the x-axis indicates different microbes (instead of chromosomal positions) and each panel represents a different taxonomical level. The y-axis indicates the p-values of the associations with faecal metabolites, which are coloured by pathway. Grey dots represent microbe-metabolite associations below the Bonferroni cut-off (adjusting for 915 metabolites \times 581 operational taxonomical units (OTUs), 382 genera, 194 families, 137 orders, 72 classes, and 34 phyla, respectively).

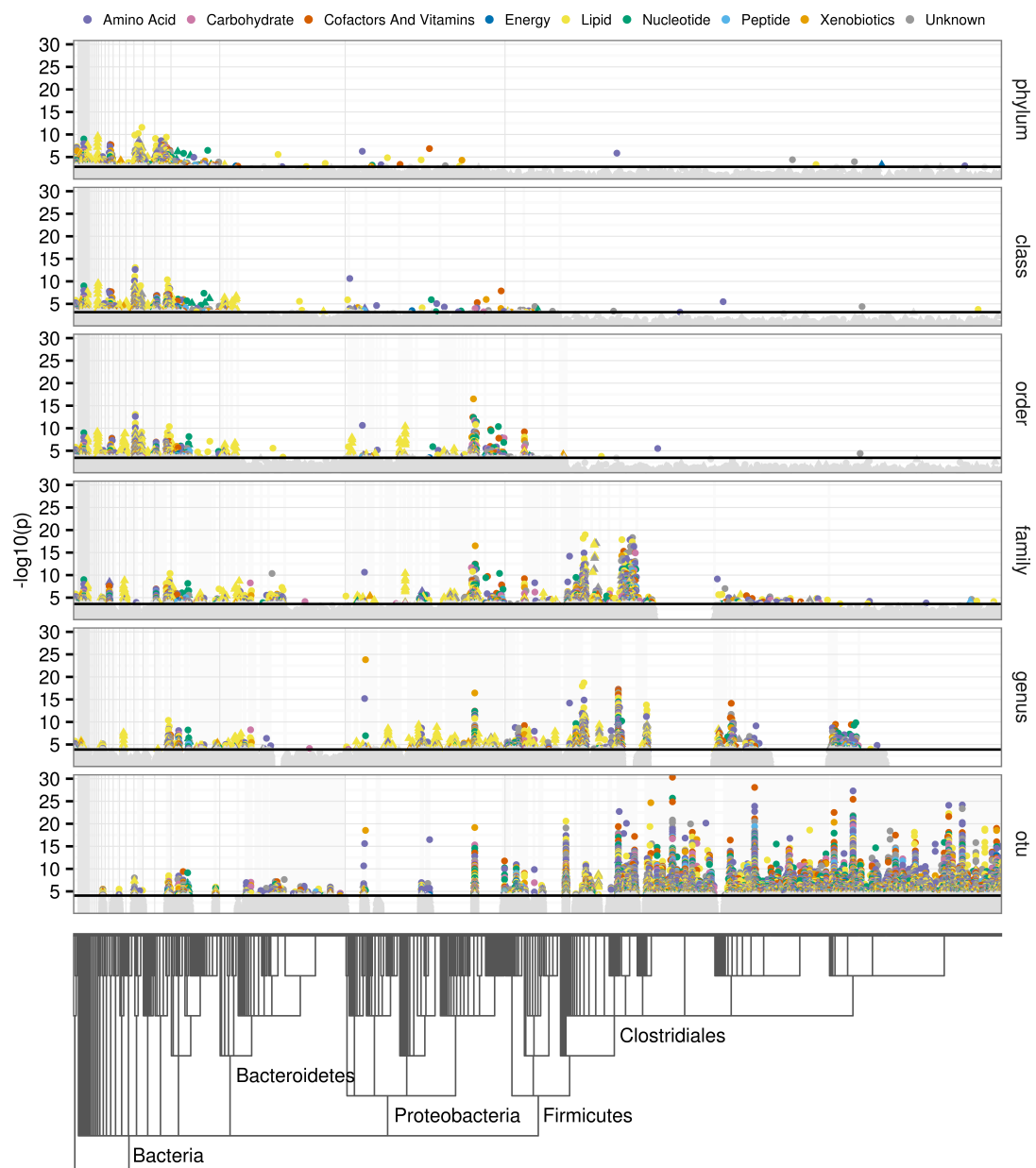
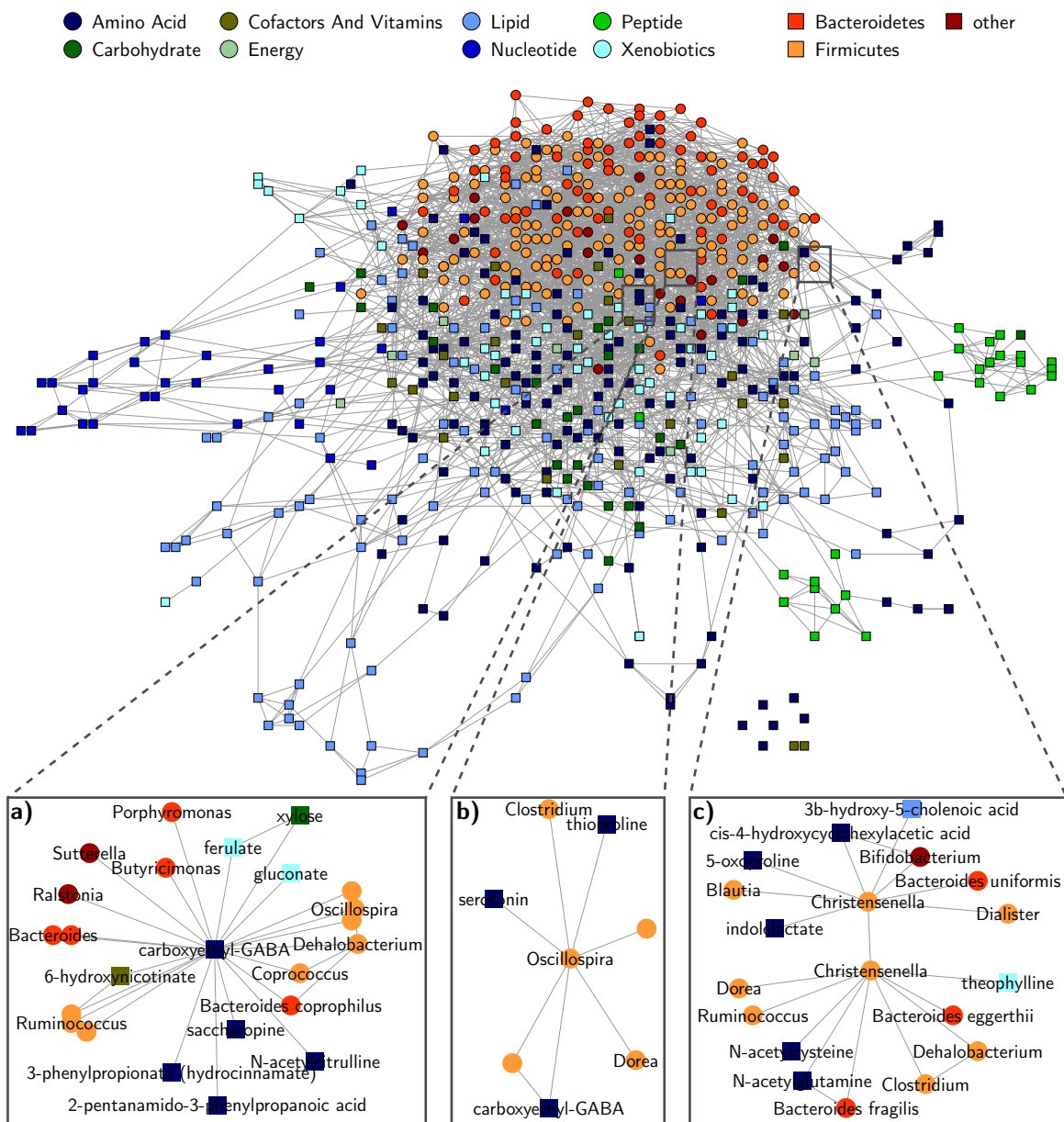


Figure 7.7 Integrating faecal metabolites and microbes. I integrated 423 faecal metabolite levels and 241 operational taxonomical units (OTUs) using a Gaussian graphical model (GGM) to explore their multivariate dependencies. These 664 variables were connected by 2460 edges. While for instance amino acids were strongly interconnected with microbes, peptides and nucleotides appeared to be more peripheral. (a) With 23 edges, 15 of them with microbes, carboxyethyl-GABA was the most densely connected node in the graph. (b) The *Oscillospira* OTU that showed the strongest association with visceral fat was directly connected with the neurotransmitter serotonin in the GGM. (c) Two *Christensenella* OTUs were included in the network. Both were independently associated with several amino acids, which might explain their association with visceral fat.



The only metabolite that was significantly different in the elderly was phytanate, which cannot be produced by humans and is thought to originate mainly from dietary intake of meat and dairy (Brink et al., 2006). Changes of phytanate levels in faeces could thus reflect dietary changes in the elderly. Phytanate is an activator of *proliferator-activated receptor α* (PPAR- α), which is involved in lipid control (Zomer et al., 2000). Due to its role in regulation of lipid metabolism (Schoonjans et al., 1996), PPAR- α is involved in several metabolic disorders, including atherosclerosis, and generally ageing (Pineda Torra et al., 1999), for instance due to response to oxidative stress (Poynter et al., 1998). Thus, decreased levels of phytanate in the elderly might lead to changes in lipid metabolism and increased susceptibility to disease.

7.3.2 The faecal metabolome is influenced by host genetics

My results show that the faecal metabolome is influenced by host genetics with 428 metabolites exhibiting a heritable component, explaining on average 17.9% ($\pm 9.7\%$) of the total variance, ranging up to 60.7%.

The host genetic influences on faecal metabolites are possibly mediated by heritable microbes. For instance, the metabolite 3-phenylpropionate is produced by several *Clostridium* species (Moss et al., 1970), which were previously shown to be heritable (Goodrich et al., 2014). Microbial production of 3-phenylpropionate might explain its heritable variance component in this data (24.3%) as well as its genome-wide significant association with the SNP rs58539483 in the lincRNA *RP11-115J23.1*, a gene that has been previously found to be associated with waist circumference (Fox et al., 2007).

On the other hand, host genetic influences might reflect efficacy of metabolic processes along with degradation and excretion of nutrients. For instance, the metabolite ratio of 5-acetylamino-6-amino-3-methyluracil and 1,3-dimethylurate, two metabolites of caffeine (Weimann et al., 2005), was strongly associated with the SNP rs35246381 close to the *N-acetyltransferase 2* (NAT2) gene. The same locus has been previously reported to be associated with other caffeine metabolites (1-methylxanthine, 4-acetamidobutanoate and 1-methylurate) in blood (Shin et al., 2014) and urine (Raffler et al., 2015). The NAT2 gene codes for a N-acetyltransferase, which catalyses the degradation of caffeine metabolites (Nyéki et al., 2003). Hence, this association likely reflects the degradation of caffeine.

7.3.3 The faecal metabolome reflects microbial composition

While faecal metabolites are moderately influenced by host genetics, they are strongly and widely associated with the gut microbiome, which explains a major proportion of the variance for most of the observed metabolites. Almost all observed OTUs and broader taxonomic groupings were significantly associated with one or more faecal metabolites. These associations reflect (i) microbial production of metabolites, such as different B vitamins (Magnúsdóttir et al., 2015), (ii) adaption of microbe composition to degradation products of the host metabolism, such as bilirubin, a metabolite of haemoglobin, which is metabolised by gut microbes (Becker et al., 2011), or (iii) microbial adaption to nutritional habits, such as fructose intake (Payne et al., 2012).

While some of the metabolites were associated with the overall composition and diversity of the microbial community, such as the B vitamins nicotinate and pantothenate or the long-chain fatty acid arachidonate, others were more specific to certain OTUs and genera, such as tyramine, a biogenic amine that is produced by the *Enterococcus* genus (De Palencia et al., 2011). This demonstrates the potential of the faecal metabolome to provide functional annotation at high resolution.

7.3.4 Faecal metabolites associate with obesity

I used the faecal metabolome to investigate potential mechanisms of microbiome-related obesity. Visceral fat mass is a measure of abdominal obesity, which was previously shown to be an independent risk factor for common complications of obesity, including CVD (Després, 2007; Fox et al., 2007).

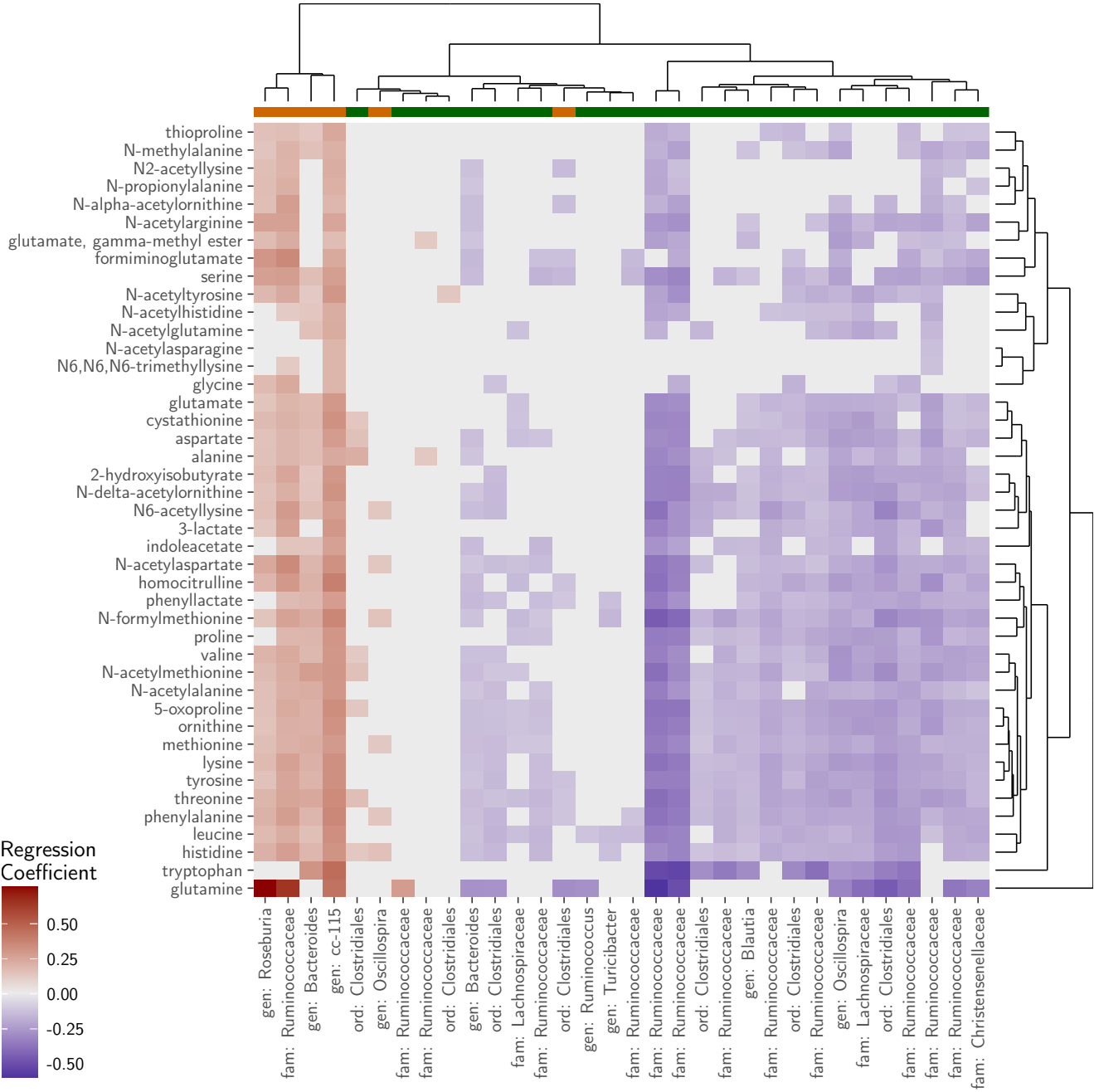
Early studies found *Firmicutes* to be overall detrimental and *Bacteroidetes* overall beneficial for obesity (Ley et al., 2006). However, these associations were disputed in a recent meta-analysis including 10 studies (Sze et al., 2016). Moreover, *Oscillospira* abundance has been associated with lower visceral fat mass (Beaumont et al., 2016), and with reduced weight gain within germ-free mice receiving human faecal transplants supplemented with *Christensenella* minute (Goodrich et al., 2014). Additionally, some *Ruminococcus* OTUs were shown to be negatively associated with visceral fat mass (Beaumont et al., 2016). Therefore, it is likely that members of the *Firmicutes* phylum have diverse functions, some of which might be protective against obesity.

Here, I found the associations of OTUs with visceral fat consistent with their metabolic profile. I identified 102 metabolites associated with visceral fat, including 43 amino acids. In the same data set – which is a subset of the data used by Beaumont et al. (2016) – 35 microbial OTUs were associated with visceral fat at a FDR of 5 %. 32 of these OTUs were associated with at least one of the visceral fat-associated amino acids. While microbial associations were not consistent within phyla, particularly not within the *Firmicutes*, all OTUs associated with increased visceral fat were strongly associated with increased abundance of amino acids and vice versa (Figure 7.8). While microbes of one taxonomical group, such as phyla but also finer groups, clearly can have very different effects on human health, I found similar associations with obesity of microbes with similar metabolic profiles. This demonstrates the potential of faecal metabolomics to complement sequencing based profiling methods.

7.3.5 Faecal metabolites associated with renal function

In patients with renal failure a considerable amount of creatinine is thought to be excreted in the gut (Jones et al., 1974), where it is further catabolised by gut microbes. 1-methylguanidine, which I found negatively correlated with renal function, is one of the microbial products of creatinine (Eyck et al., 1968). Thus, the observed increase in faecal concentration of 1-methylguanidine with decline of renal function is most likely a consequence of insufficient renal excretion of creatinine. However, 1-methylguanidine was also found to be toxic and, for instance, increases mortality in rats (Yokozawa et al., 1989). Consequently, the decrease of renal function, the

Figure 7.8 Association of faecal amino acids with visceral fat. 43 faecal amino acids significantly positively associated with visceral fat. Several microbes have been previously identified to be associated with obesity and visceral fat. Here, I found 35 microbial operational taxonomical units (OTUs) associated with visceral fat at a FDR of 5%, 7 positively (orange) and 28 negatively (green). 32 of these OTUs were also associated with at least one of the visceral fat-associated amino acids. Red tiles indicate positive associations of microbes and metabolites ($\beta > 0$) and blue ones negative associations ($\beta < 0$); grey tiles indicate non-significant associations (FDR>5%). Clearly, microbes that are positively associated with visceral fat are also positively associated with the abundance of amino acids and vice versa, indicating that their metabolic activity is more closely related to human health than their taxonomy.



accumulation of 1-methylguanidine, and gut dysbiosis might mutually enforce each other (Ramezani et al., 2016). However, here we analysed a small set of relatively renal-healthy individuals and more in-depth studies are needed to fully elucidate these complex relationships.

7.4 Conclusion

Here, I analysed metabolomics measurements from faecal samples to investigate their potential use as biomarkers of ageing. While ageing and renal function appear to be only moderately associated with faecal metabolites, I found obesity strongly associated with many metabolites. These associations indicate that increased abundance of amino acids in the gut, which are potentially produced by microbes, might mediate the association of abdominal obesity and changes of microbial composition.

I show that the faecal metabolome, while moderately influenced by host genetics, largely reflects the microbial community and thus indeed provides a functional readout of microbial metabolism. As metabolites are the main means of communication between the host and (gut) microbiome, the faecal metabolome can be used as an intermediate phenotype that promotes microbial effects on the host. This might facilitate future studies to overcome the limitations of sequencing based methods, most importantly the lack of functional annotation.

However, I analysed observational data only and cannot conclude causality for any of the observed correlations. Also, due to the novelty of faecal metabolomics, the study lacks independent replication and the stability of the observed associations across populations and time has yet to be addressed. Moreover, I focused on the relationship of faecal metabolites with host and microbial genetics, however, future studies should investigate the influence of environmental factors, most importantly diet. Lastly, there are limitations regarding the collection of faecal samples, which are stored by participants until they arrive in the study centre. This storage has a considerable effect on the microbes (Choo et al., 2015) and on the metabolome.

Summarising, faecal metabolomics appears to be a useful tool to complement future microbiome studies with functional annotation and hence promote knowledge on mechanisms of host-microbe interactions and their impact on human health.

CHAPTER 8

Differential multi-fluid networks identify processes involved in end-stage renal disease

The aim of this chapter was to identify metabolic processes that are affected by end-stage renal disease (ESRD). To obtain a more holistic image of the metabolic processes associated with ESRD, I modelled metabolic networks spanning three fluids – plasma, urine, and saliva – using Gaussian graphical models (GGMs). Differences of these cross-fluid metabolic networks between cases and controls were investigated using permutation testing. This yielded three metabolic pathways that are disrupted in ESRD patients, independently of other metabolic changes.

Previous studies demonstrated the tremendous impact of renal function on the metabolism in various fluids (Barrios et al., 2016). In chapter 6, I investigated the association of circulating metabolites with renal function and found strong associations of eGFR with lipid composition, amino acids, and energy metabolism. Also, Sekula et al. (2016) identified more than 50 circulating metabolites, measured on the Metabolon platform, associated with renal function in the KORA and TwinsUK cohorts. While these were the largest studies investigating the healthy spectrum of renal disease, several smaller studies investigating changes of circulating metabolites in (end stage) renal disease patients found – amongst others – altered abundance of amino acids, uremic solutes and acylcarnitines (Sun et al., 2012; Niewczasz et al., 2014; Toyohara et al., 2010).

With urine being the output of renal filtration and also being one of the most accessible bio-fluids, several studies aimed to identify urinary biomarkers of renal function. NMR spectroscopy helped to identify several candidate biomarkers, including citrate and threonine (Posada-Ayala et al., 2014), which were later replicated by an independent study (Pallet et al., 2014). Another study found decreased excretion of citrulline and asymmetric dimethylarginine (ADMA) in CKD patients (Duranton et al., 2014).

As patients with ESRD might not have urine, saliva is a possible alternative, which is as accessible as urine. Also, metabolic changes in urine might reflect other physiological changes than the decline of renal function, for instance of the genitourinary system (Issaq et al., 2008; Kind et al., 2016). Hence, several studies investigated salivary changes in kidney disease patients and found marked differences in concentrations of creatinine and urea (Lasisi et

al., 2016), electrolytes (Tomás et al., 2008) and phosphorus (Savica et al., 2008), however, no comprehensive analysis of the association between the salivary metabolome and renal function has been conducted, yet.

However, studies so far have largely been focusing on a single fluid, analysing the bivariate associations of individual metabolite levels with renal function, thus ignoring the interdependencies of metabolites across fluids, which are connected through metabolic reactions and transport processes (see section 1.3). As outlined in section 1.3.3.2, GGMs have been shown to infer cross-fluid metabolic processes from observational data (Krumstiek et al., 2011; Do et al., 2015). Rather than analysing changes of individual metabolites, GGMs facilitate the analysis of metabolic reactions and modules, and thus to investigate the driving factors behind the metabolic disturbances that are commonly observed in ESRD. Analysing differences of such metabolic networks between conditions can help to formulate hypothesis about the disease processes (Valcárcel et al., 2011).

Here, I aimed to identify metabolic differences in plasma, urine and saliva samples of 72 advanced CKD patients from the GANI_MED cohort compared to 906 controls from the SHIP cohort, representing the general population. To identify metabolic processes that potentially account for the substantial metabolic shift, I used GGMs to model cross-fluid metabolic pathways in cases and controls and identified differences thereof by permutation testing as suggested before (Valcárcel et al., 2011). This is the first metabolic analysis of CKD combining three different fluids as well as the first application of differential networks in renal research.

8.1 Methods

8.1.1 Study populations

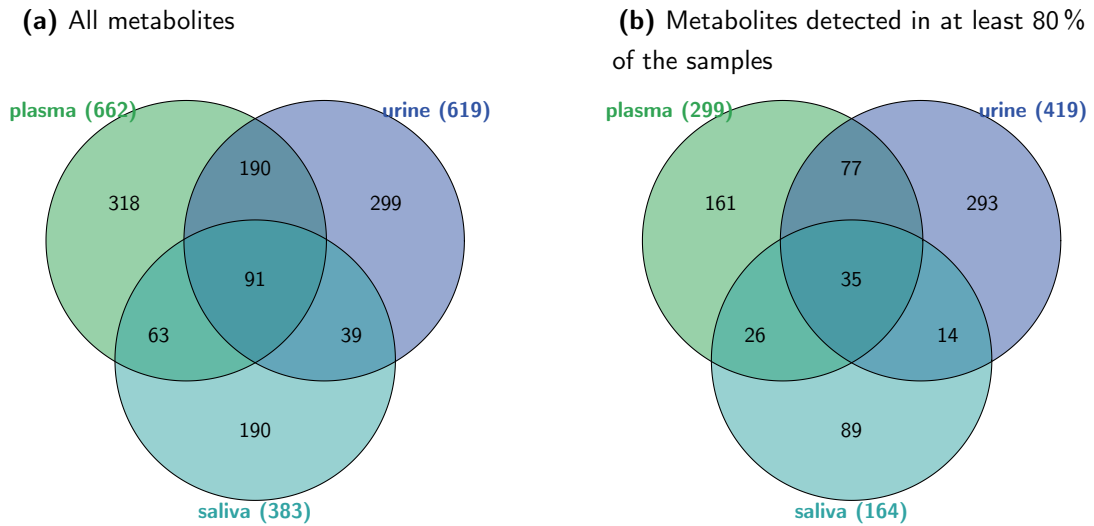
In this chapter, I analysed data from the SHIP (Völzke et al., 2011) and GANI_MED (Grabe et al., 2014) cohorts.

72 CKD patients from the GANI_MED renal cohort (section 3.1.3) with metabolomics data available were used as cases, and 906 individuals from the SHIP (section 3.1.3) study served as controls.

8.1.2 Metabolomics measurements and normalisation

Metabolon's untargeted V3 platform (section 3.3.1) was used to profile plasma, urine, and saliva samples from all 978 individuals. Cases were randomised across run-days. 662 metabolites were detected in plasma, 619 in urine and 383 in saliva (Figure 8.1a); 299, 419 and 164 of these were detected in at least 80 % of the study population (Figure 8.1b). Measurements were scaled by run-day medians as described before, however, run-day medians were calculated using the control individuals from the SHIP cohort only to avoid bias due to the sometimes dramatically different metabolite levels in cases.

Figure 8.1 Metabolites detected in plasma, urine and saliva. (a) Metabolites were measured using an untargeted mass spectrometry (MS) platform in plasma, urine and saliva. (b) More than the half of the metabolites were detected in more than 80 % of the samples.



Metabolite concentrations in urine and saliva strongly depend on the dilution with water, which reduces the total amount of metabolites. To correct for the dilution effect, many studies scale measurements relative to the concentration of creatinine (Vogl et al., 2016), however this is problematic when analysing renal function. Thus, I used probabilistic quotient normalisation to account for dilution (Dieterle et al., 2006) in urine and saliva samples. To this end, I estimated a reference sample as median of all metabolites that were detected in all samples. Then, I calculated the deviation from this reference for each sample as median of all metabolite fold-changes. This deviance from the reference was used as dilution factor and all metabolite levels of the respective sample were scaled by it. As for the run-day normalisation, the reference sample was calculated from controls only, individually for urine and saliva measurements. All metabolite concentrations were subsequently log-transformed and scaled to mean 0 and a SD 1.

As network inference relies on full data matrices, I excluded metabolites with more than 20 % missing values in the combined dataset of cases and controls. The remaining missing values were imputed using the R package mice (see section 3.4.5). Case and control groups differed with respect to age, sex, and BMI (Table 8.1). Hence, I adjusted for these factors by fitting linear models for each metabolite. Residuals of these models, scaled to mean 0 and SD of 1, were used for further analyses.

8.1.3 Statistical analysis

I started with calculating logistic regression models (see section 3.4.2) for each of the 882 metabolites using the residuals adjusted for age, sex, and BMI, to identify metabolites differing between cases and controls. Results were adjusted for multiple testing using Bonferroni correction for 882 tests (see section 3.4.2.1). To identify pathways that were jointly affected by the disease, I tested for pathway enrichment as described in section 3.4.3.

Individually for cases and controls, I inferred GGMs to model cross-fluid metabolic processes using the GeneNet algorithm (Schäfer et al., 2005) as described in section 3.4.7. I normalised each metabolite to mean 0 and standard deviation 1 individually for cases and controls prior to network inference, to avoid misleading correlations of metabolites in the subsequent permutation testing. For each edge the FDR was estimated and edges passing the FDR-threshold of 5 % were included in the respective GGMs.

To test for significant differences between the networks of cases and controls, I used a permutation based approach. To this end, I combined the datasets for cases and controls and permuted the class labels (case/control) 10,000,000 times. In each permutation, I divided the dataset in two sets (of 906 and 72 samples, respectively), inferred a GGMs for both, and calculated the differences of the resulting shrinkage partial correlations. Consequently, background distributions of shrinkage partial correlation differences were obtained for each edge. These were used to calculate empirical two-sided p-values, assessing the significance of the differences of shrinkage partial correlations between the case and the control networks.

Empirical p-values were then adjusted for multiple comparisons. To reduce the burden of multiple testing (there are $388,521 = 882 \times (882 - 1)/2$ potential edges in the network), I only tested edges for significant differential (shrinkage partial) correlations of edges that were included in either (or both) of the networks for cases and controls. Consequently, the empirical p-values for edge differences were adjusted for 5937 tests using Bonferroni correction ($p < 8.4 \times 10^{-6} = 0.05/5937$).

I used the control GGM as reference graph to compute a layout for visualisation, as it is meant to represent the healthy metabolic network. Measures of graph topology and node centrality were calculated for both networks as described in section 3.4.9.

Table 8.1 Population characteristics.

	SHIP (controls)	GANI_MED (cases)
<i>N</i>	906	72
<i>Age, years</i>	49.94±13.64	64.40±15.38
<i>Gender (female)</i>	512 (56.5 %)	22 (30.6 %)
<i>BMI, kg/m²</i>	27.35±4.50	30.17±6.79
<i>Dialysis</i>	0	44 (61.1 %)

8.2 Results

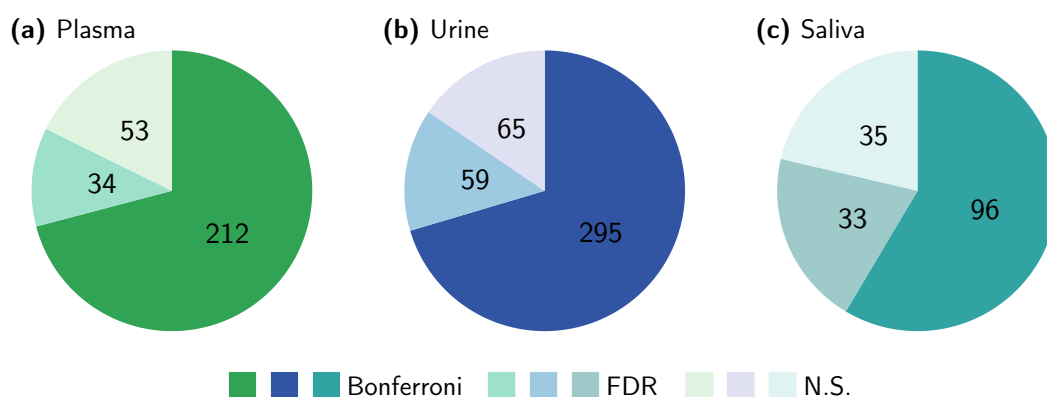
I analysed metabolomics data from plasma, urine, and saliva from 72 CKD patients from the GANL_MED cohort, by comparing them with 906 individuals from the SHIP cohort, which represent the general population (Table 8.1). The metabolic profiles of CKD patients were immensely different from the general population. Principal component analysis (PCA) shows that the disease accounts for a major proportion of the observed variance in each of the three fluids, independently of age, sex and BMI (Figure E.2).

8.2.1 Univariate differences between cases and controls

603 metabolites (68 %) were significantly different between CKD cases and controls after Bonferroni correction for multiple testing ($p < 5.7 \times 10^{-5} = 0.05/882$ metabolites). 212 of these were measured in plasma, 295 in urine, and 96 in saliva (Figure 8.2, Table E.1). Amino acids (enrichment $p = 1.2 \times 10^{-3}$) and nucleotides ($p = 1.0 \times 10^{-3}$) were significantly enriched in plasma of CKD patients, while nucleotides were significantly depleted in urine ($p < 1.0 \times 10^{-4}$) of cases compared to controls. In contrast, lipids were depleted in both plasma ($p < 1.0 \times 10^{-4}$) and urine ($p = 1.0 \times 10^{-4}$) of renal disease patients. Particularly lysolipids were found at lower concentrations in patients' blood ($p < 1.0 \times 10^{-4}$).

Seven metabolites were significantly different between the CKD patients and controls across all three fluids: four amino acids – N-acetylphenylalanine, acisoga, 3-methyl-2-oxovalerate, alpha-hydroxyisovalerate – the lipid azelate, citrate and the vitamin pantothenate (Figure E.3). Moreover, 74 metabolites were associated with CKD status in two fluids: 18 of them in plasma and saliva, 5 in saliva and urine, and the remaining 51 metabolites in plasma and urine.

Figure 8.2 Metabolites differing between CKD cases and controls. I found 603 of the 882 analysed metabolites significantly different between cases and controls after stringent correction for multiple testing (Bonferroni). Additionally, 126 metabolites were significantly different at a FDR of 5 %.



8.2.2 Multi-fluid graphical models

In the next step, I aimed to further investigate the pathways affected by kidney disease and to identify the driving factors underlying these profound metabolic differences. To this end, I inferred a GGM from the controls as a model of the healthy metabolic network. 4404 edges passed the significance threshold of $FDR < 5\%$, 972 of them between different fluids (Table 8.2). The distribution of node degrees approximated the power law (Figure 8.3), thus forming a scale-free network. Shrinkage partial correlations favour sparse networks while excluding spurious (mediated) correlations. Thus, the GGM was much sparser than a CN based on ordinary correlations, which consisted of 90,982 (24.4% of all edges) significant edges. 46% of the tuples (117) and triplets (35) (Figure 8.1b) of the same metabolite measured in different fluids were connected in the final GGM.

The metabolic model of cases only contained 2509 edges. Despite the much fewer edges compared to the controls GGM, the distribution of node degrees still followed the power law. However, while in the controls network all 882 metabolites were connected with each other in one large component, 52 individual metabolites and two pairs of metabolites were separated from the remaining 826 metabolites, which formed the largest component in the case graph. Both graphs showed small world properties, with high degrees of clustering but small shortest path lengths (Table 8.3).

8.2.3 Network differences

Finally, I tested for significant differences between the graphs for cases and controls by pooling the datasets, permuting class labels, and estimating background distributions of differential shrinkage partial correlations. 86 edges were significantly different between the two GGMs after adjustment for multiple testing using Bonferroni correction ($p < 8.4 \times 10^{-6} = 0.05/5937$ tested edges) (Table 8.4 and Figure 8.4).

Three of these differential edges link the same metabolite between different fluids: 7-methylxanthine in plasma and urine (cases GGM: $p = 1.2 \times 10^{-13}$, controls GGM: $p = 3.3 \times 10^{-3}$, difference: $p < 1.0 \times 10^{-7}$), caffeine in plasma and saliva (cases GGM: $p = 1.3 \times 10^{-10}$, controls GGM: $p = 5.6 \times 10^{-8}$, difference: $p = 2.0 \times 10^{-7}$), and cortisol in plasma and urine (cases GGM: $p = 1.3 \times 10^{-1}$, controls GGM: $p = 2.2 \times 10^{-16}$, difference: $p = 4.0 \times 10^{-7}$). Additionally, 27 differential edges were cross-fluid, including some closely related metabolites such as pregn steroid monosulfate in plasma with pregnen-diol disulfate in urine (cases GGM: $p = 7.7 \times 10^{-1}$, controls GGM: $p = 1.9 \times 10^{-5}$, difference: $p < 1.0 \times 10^{-7}$).

The remaining 55 differential edges connected metabolites within the same fluid, 13 of them in plasma, 23 in urine, and 19 in saliva (Table 8.2c). 2, 9, and 12, respectively, of those connect metabolites of the same general group.

Table 8.2 Cross-fluid edges. (a) The Gaussian graphical model (GGM) for the SHIP cohort (controls) consisted of 4404 edges connecting the 882 metabolites. 936 of these edges were between fluids. (b) Due to the much smaller sample size the network for cases consisted of 2509 edges. (c) 86 edges were significantly different between the models.

(a) Controls				(b) Cases				(c) Differential			
	plasma	saliva	urine	plasma	saliva	urine		plasma	saliva	urine	
plasma	1069 (24.3 %)	147 (3.3 %)	637 (15.3 %)	570 (22.7 %)	99 (3.9 %)	354 (14.1 %)		13 (15.1 %)	5 (5.8 %)	16 (18.6 %)	
saliva		699 (15.2 %)	152 (3.5 %)		489 (19.5 %)	169 (6.7 %)			19 (22.1 %)	10 (11.6 %)	
urine			1694 (38.5 %)			828 (33.0 %)				23 (26.7 %)	

Figure 8.3 Network properties. Gaussian graphical models (GGMs) were inferred individually for cases (red) and controls (yellow). (a) The degree distribution shows the cumulative proportion of nodes with a certain number of edges connected to them. The total number of edges, and consequently the average degree, is lower in the cases network due to the much smaller number of samples. (b) The betweenness centrality of a node indicates the percentage of pairwise shortest paths that contain this node.

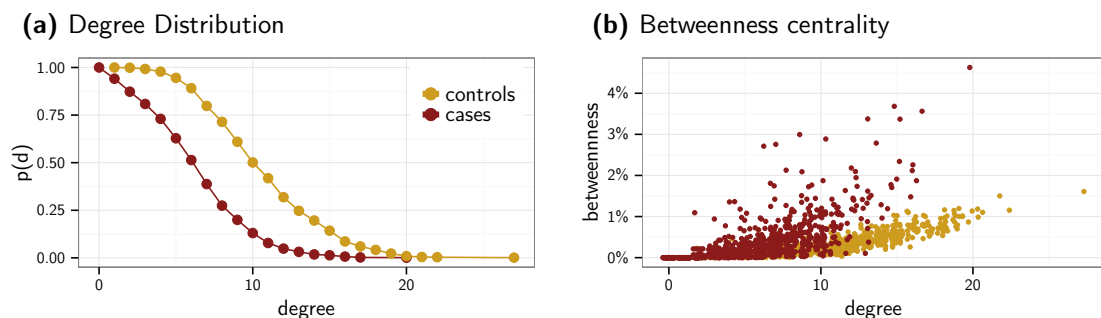


Table 8.3 Properties of the multi-fluid GGMs. Topological features were calculated for both networks, inferred from data of cases and controls. The network for cases has less edges and consequently a lower average node degree and longer path lengths, due to the much smaller number of cases. However, in relation to the high degree of local and global clustering both graphs have short shortest path length and hence are small-world networks.

	Controls	Cases
# Edges	4404	2509
Average Degree	10.0 (± 3.9)	5.7 (± 3.4)
Global Clustering	10.9 %	21.9 %
Local Clustering	18.4 (± 18.9) %	28.1 (± 26.9) %
Average Betweenness	0.282 (± 0.255) %	0.362 (± 0.522) %
Average shortest paths	3.5 (± 0.85)	4.6 (± 1.4)
Small worldness	14.8	38.7

Table 8.4 Differential edges. 86 edges were significantly different in the metabolic network of chronic kidney disease (CKD) cases compared to controls. For each edge the shrinkage partial correlations, p-values, and edge-betweenness (B) from the graphs of controls and cases, respectively, are shown. Empirical p-values were computed from 10,000,000 random samples to assess the significance of the differences between the networks.

Node 1		Node 2		Controls			Cases			differential p
Fluid	Metabolite	Fluid	Metabolite	p-cor	p-value	B	p-cor	p-value	B	
plasma	androsterone sulfate	plasma	S-methylcysteine	0.01	7.8×10^{-1}		0.04	2.3×10^{-7}	0.12%	$<1.0 \times 10^{-7}$
plasma	EDTA	plasma	X - 17628	0.18	2.2×10^{-16}	0.04%	0.05	6.1×10^{-10}	0.01%	$<1.0 \times 10^{-7}$
plasma	pregn steroid monosulfate	urine	pregnen-diol disulfate	0.08	1.9×10^{-5}	0.01%	0.00	7.7×10^{-1}		$<1.0 \times 10^{-7}$
plasma	urea	plasma	X - 11727	0.31	2.2×10^{-16}	0.03%	-0.01	4.6×10^{-1}		$<1.0 \times 10^{-7}$
plasma	7-methylxanthine	urine	7-methylxanthine	0.06	3.3×10^{-3}		0.06	1.2×10^{-13}	0.04%	$<1.0 \times 10^{-7}$
saliva	leucine	saliva	isoleucine	0.23	2.2×10^{-16}	0.02%	0.08	2.2×10^{-16}	0.05%	$<1.0 \times 10^{-7}$
saliva	leucine	saliva	tyrosine	0.14	5.7×10^{-14}	0.03%	-0.01	5.2×10^{-1}		$<1.0 \times 10^{-7}$
saliva	phenylalanine	saliva	isoleucine	0.12	2.0×10^{-10}	0.01%	0.06	2.6×10^{-13}	0.05%	$<1.0 \times 10^{-7}$
saliva	trans-urocanate	urine	X - 12695	0.01	7.8×10^{-1}		-0.05	1.3×10^{-11}	0.18%	$<1.0 \times 10^{-7}$
saliva	isoleucine	saliva	valine	0.20	2.2×10^{-16}	0.03%	0.08	2.2×10^{-16}	0.04%	$<1.0 \times 10^{-7}$
saliva	X - 11854	saliva	X - 13671	-0.02	3.2×10^{-1}		0.03	8.5×10^{-5}	0.07%	$<1.0 \times 10^{-7}$
saliva	X - 11854	saliva	X - 14081	0.27	2.2×10^{-16}	0.00%	0.08	2.2×10^{-16}	0.02%	$<1.0 \times 10^{-7}$
saliva	X - 12776	urine	tigloylglycine	0.01	7.3×10^{-1}		-0.03	1.4×10^{-4}	0.24%	$<1.0 \times 10^{-7}$
saliva	X - 14081	saliva	X - 16612	0.17	2.2×10^{-16}	0.03%	0.07	2.2×10^{-16}	0.05%	$<1.0 \times 10^{-7}$
saliva	X - 14196	saliva	N-acetylserine	0.07	2.3×10^{-4}	0.11%	-0.02	2.9×10^{-3}		$<1.0 \times 10^{-7}$
saliva	X - 14196	saliva	X - 19496	0.01	7.8×10^{-1}		0.03	9.9×10^{-6}	0.17%	$<1.0 \times 10^{-7}$
saliva	X - 19807	urine	3,7-dimethylurate	0.01	6.9×10^{-1}		0.03	7.4×10^{-5}	0.25%	$<1.0 \times 10^{-7}$
urine	X - 12329	urine	X - 17185	0.12	3.9×10^{-10}	0.06%	0.00	9.3×10^{-1}		$<1.0 \times 10^{-7}$
plasma	caffeine	saliva	caffeine	0.10	5.6×10^{-8}	0.13%	0.05	1.3×10^{-10}	0.13%	2.0×10^{-7}
plasma	EDTA	plasma	X - 17629	0.11	4.2×10^{-9}	0.06%	0.05	1.8×10^{-9}	0.02%	2.0×10^{-7}
plasma	X - 12798	urine	3-hydroxyphenylacetate	-0.04	2.3×10^{-2}		0.04	1.2×10^{-7}	0.18%	2.0×10^{-7}
saliva	leucine	saliva	valine	0.20	2.2×10^{-16}	0.03%	0.07	2.2×10^{-16}	0.06%	2.0×10^{-7}
saliva	caproate (6:0)	saliva	valerate	0.34	2.2×10^{-16}	0.05%	0.09	2.2×10^{-16}	0.04%	2.0×10^{-7}
saliva	X - 13671	saliva	X - 14081	0.02	4.2×10^{-1}		0.03	1.5×10^{-5}	0.07%	2.0×10^{-7}
urine	cortisone	urine	alpha-CEHC glucuronide	-0.03	1.1×10^{-1}		0.03	2.2×10^{-4}	0.19%	2.0×10^{-7}

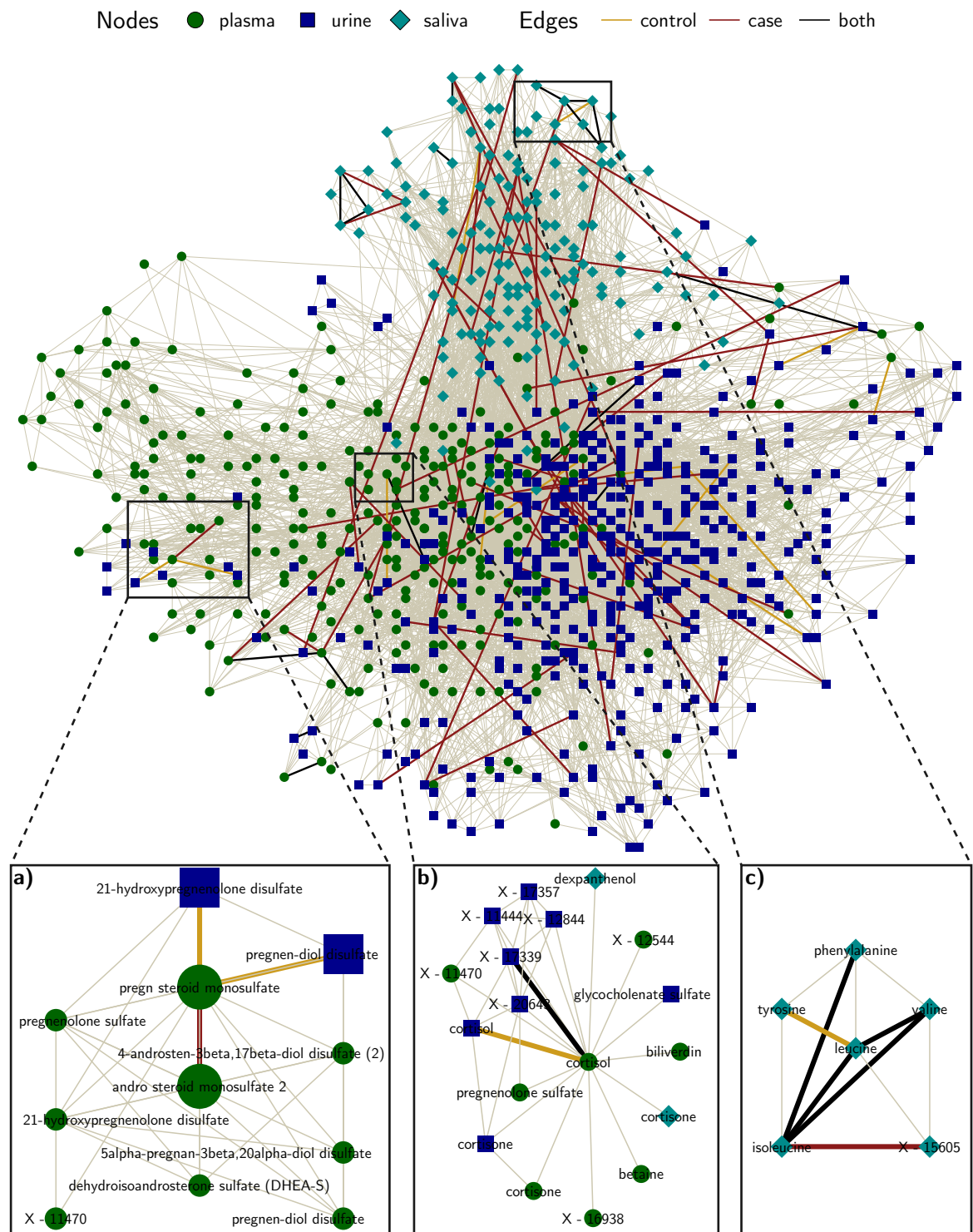
continued on next page . . .

plasma	paraxanthine	urine	1-methylxanthine	0.10	2.8×10^{-8}	0.04%	0.01	2.0×10^{-1}	2.0×10^{-7}
urine	glutamine	urine	4-vinylphenol sulfate	-0.02	4.1×10^{-1}		-0.04	8.5×10^{-7}	0.27% 2.0×10^{-7}
urine	7-methylxanthine	urine	1-methylurate	0.16	2.2×10^{-16}	0.03%	0.02	2.9×10^{-2}	2.0×10^{-7}
plasma	cortisol	urine	cortisol	0.16	2.2×10^{-16}	0.03%	-0.01	1.3×10^{-1}	4.0×10^{-7}
plasma	1-palmitoylglycerophosphoinositol	urine	theobromine	-0.02	2.7×10^{-1}		0.03	2.4×10^{-5}	0.20% 4.0×10^{-7}
saliva	X - 19807	urine	3-methyluracil	-0.01	6.0×10^{-1}		-0.03	1.1×10^{-4}	0.22% 4.0×10^{-7}
urine	cinnamoylglycine	urine	X - 17313	0.03	7.9×10^{-2}		-0.03	1.6×10^{-5}	0.28% 4.0×10^{-7}
urine	cis-aconitate	urine	X - 17736	0.02	4.2×10^{-1}		-0.03	3.6×10^{-5}	0.30% 6.0×10^{-7}
urine	X - 12123	urine	X - 17736	0.02	2.2×10^{-1}		-0.03	1.9×10^{-5}	0.31% 6.0×10^{-7}
saliva	adenine	saliva	trans-urocanate	-0.02	2.5×10^{-1}		0.05	1.2×10^{-9}	0.12% 1.0×10^{-6}
saliva	trans-urocanate	urine	X - 18943	-0.04	2.2×10^{-2}		0.04	2.9×10^{-6}	0.11% 1.0×10^{-6}
plasma	EDTA	plasma	HWESASLLR	-0.02	3.2×10^{-1}		-0.04	4.5×10^{-7}	0.04% 1.0×10^{-6}
urine	cortisone	urine	X - 15636	0.02	3.8×10^{-1}		-0.04	1.7×10^{-8}	0.22% 1.0×10^{-6}
plasma	glutamine	urine	3-hydroxyphenylacetate	-0.02	3.7×10^{-1}		0.04	1.1×10^{-7}	0.26% 1.2×10^{-6}
plasma	pelargonate (9:0)	urine	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	-0.04	4.5×10^{-2}		0.04	1.2×10^{-6}	0.44% 1.2×10^{-6}
saliva	histidine	urine	X - 12844	-0.02	3.1×10^{-1}		0.03	3.7×10^{-5}	0.35% 1.2×10^{-6}
urine	X - 12258	urine	X - 17313	0.01	6.7×10^{-1}		0.03	2.2×10^{-4}	0.08% 1.4×10^{-6}
urine	X - 13462	urine	X - 16394	0.20	2.2×10^{-16}	0.09%	0.07	2.2×10^{-16}	0.11% 1.4×10^{-6}
plasma	EDTA	plasma	lactate	0.03	1.1×10^{-1}		-0.03	1.1×10^{-5}	0.19% 1.6×10^{-6}
saliva	vanillin	saliva	X - 19489	0.12	1.0×10^{-10}	0.03%	0.05	2.8×10^{-9}	0.01% 1.6×10^{-6}
urine	cis-aconitate	urine	X - 14951	-0.05	1.2×10^{-2}		0.05	1.8×10^{-10}	0.20% 1.6×10^{-6}
plasma	p-cresol sulfate	plasma	4-methylcatechol sulfate	0.11	5.6×10^{-9}	0.17%	0.06	6.7×10^{-13}	0.04% 1.8×10^{-6}
plasma	DSGEGDFAEGGGVR	plasma	X - 17629	0.04	2.0×10^{-2}		-0.03	1.5×10^{-5}	0.03% 2.0×10^{-6}
urine	X - 17320	urine	S-(3-hydroxypropyl)mercapturic acid (HPMA)	0.08	1.2×10^{-5}	0.10%	-0.01	2.6×10^{-1}	2.0×10^{-6}
saliva	X - 11854	saliva	X - 16612	0.16	2.2×10^{-16}	0.03%	0.07	2.2×10^{-16}	0.03% 2.0×10^{-6}
urine	p-cresol sulfate	urine	4-methylcatechol sulfate	0.07	1.2×10^{-4}	0.26%	0.06	2.3×10^{-12}	0.08% 2.2×10^{-6}
plasma	creatine	plasma	pyroglutamine	-0.20	2.2×10^{-16}	0.05%	0.01	3.0×10^{-1}	2.4×10^{-6}
plasma	pyroglutamine	urine	X - 12339	-0.04	1.8×10^{-2}		0.03	3.1×10^{-4}	0.10% 2.6×10^{-6}
urine	glutamine	urine	tyrosine	0.00	9.7×10^{-1}		0.04	4.6×10^{-6}	0.04% 2.6×10^{-6}
plasma	HWESASXX	urine	cyclo(gly-pro)	0.00	9.6×10^{-1}		-0.04	7.5×10^{-7}	0.14% 2.8×10^{-6}
plasma	cortisol	urine	X - 17339	-0.08	2.9×10^{-5}	0.07%	0.03	1.3×10^{-4}	0.10% 3.0×10^{-6}

continued on next page . . .

saliva	X - 12776	urine	cortisone	−0.02	2.0×10^{-1}		−0.04	6.8×10^{-6}	0.25%	3.2×10^{-6}
saliva	citrulline	saliva	2-hydroxyglutarate	0.04	5.2×10^{-2}		0.03	1.5×10^{-4}	0.07%	3.2×10^{-6}
urine	xanthine	urine	X - 12026	0.11	4.9×10^{-9}	0.04%	−0.01	2.6×10^{-1}		3.4×10^{-6}
saliva	X - 12776	saliva	docosadioate	−0.02	2.9×10^{-1}		0.03	9.1×10^{-5}	0.25%	3.6×10^{-6}
plasma	pregn steroid monosulfate	urine	21-hydroxypregnenolone disulfate	0.08	1.0×10^{-5}	0.02%	0.01	5.1×10^{-1}		3.8×10^{-6}
urine	1,7-dimethylurate	urine	X - 18838	0.01	4.6×10^{-1}		0.03	5.2×10^{-5}	0.29%	3.8×10^{-6}
plasma	Isobar: glucose, fructose, mannose, galactose, allose, altrose, etc.	plasma	X - 11727	0.37	2.2×10^{-16}	0.04%	0.09	2.2×10^{-16}	0.07%	4.0×10^{-6}
plasma	Isobar: glucose, fructose, mannose, galactose, allose, altrose, etc.	plasma	X - 12776	−0.04	5.7×10^{-2}		−0.06	5.6×10^{-13}	0.26%	4.2×10^{-6}
plasma	X - 11529	saliva	p-cresol sulfate	−0.04	2.6×10^{-2}		0.04	6.2×10^{-7}	0.09%	4.4×10^{-6}
urine	X - 12216	urine	X - 12687	0.17	2.2×10^{-16}	0.07%	0.06	2.6×10^{-14}	0.15%	4.4×10^{-6}
plasma	HWESASXX	saliva	X - 19852	−0.01	5.3×10^{-1}		−0.04	2.8×10^{-6}	0.12%	4.6×10^{-6}
urine	3-methylglutarate	urine	alpha-CEHC sulfate	0.01	6.4×10^{-1}		0.05	4.3×10^{-9}	0.06%	4.8×10^{-6}
urine	phenylcarnitine	urine	O-sulfo-L-tyrosine	−0.01	6.2×10^{-1}		−0.04	2.8×10^{-6}	0.24%	5.0×10^{-6}
saliva	X - 12803	saliva	4-hydroxybutyrate (GHB)	0.06	1.2×10^{-3}		0.04	1.2×10^{-7}	0.35%	5.2×10^{-6}
saliva	dexpanthenol	urine	X - 12680	0.05	1.6×10^{-2}		−0.04	1.7×10^{-8}	0.36%	5.2×10^{-6}
plasma	N-methyl proline	saliva	valerate	−0.01	7.4×10^{-1}		0.04	7.7×10^{-6}	0.16%	5.4×10^{-6}
plasma	theobromine	saliva	3-phenylpropionate (hydrocinnamate)	0.00	8.9×10^{-1}		−0.03	2.4×10^{-4}	0.24%	5.6×10^{-6}
saliva	lysine	urine	gentisate	−0.03	9.8×10^{-2}		0.03	2.3×10^{-5}	0.15%	5.8×10^{-6}
plasma	pelargonate (9:0)	urine	X - 12846	0.01	7.1×10^{-1}		−0.03	1.5×10^{-5}	0.29%	6.0×10^{-6}
saliva	isoleucine	saliva	X - 15605	0.01	5.7×10^{-1}		−0.04	5.1×10^{-6}	0.10%	6.2×10^{-6}
plasma	nicotinamide	urine	X - 16947	−0.03	1.3×10^{-1}		−0.06	3.6×10^{-15}	0.31%	6.4×10^{-6}
urine	4-hydroxyhippurate	urine	X - 20588	0.01	7.8×10^{-1}		0.03	2.2×10^{-4}	0.22%	6.4×10^{-6}
saliva	valerate	urine	hypoxanthine	0.01	7.0×10^{-1}		0.03	2.8×10^{-4}	0.12%	6.8×10^{-6}
urine	cis-aconitate	urine	alpha-hydroxyisovalerate	0.00	8.4×10^{-1}		0.04	7.6×10^{-7}	0.26%	7.2×10^{-6}
urine	X - 12687	urine	stachydrine	−0.02	3.9×10^{-1}		−0.03	2.9×10^{-4}	0.13%	7.4×10^{-6}
plasma	azelate (nonanedioate)	plasma	1-methyl-2-piperidinecarboxylic acid	0.01	6.9×10^{-1}		0.04	1.9×10^{-6}	0.13%	8.0×10^{-6}
plasma	pregn steroid monosulfate	plasma	andro steroid monosulfate 2	−0.02	2.6×10^{-1}		0.04	4.7×10^{-7}	0.04%	8.0×10^{-6}
plasma	N-delta-acetylornithine	urine	3,7-dimethylurate	−0.00	8.9×10^{-1}		−0.03	1.4×10^{-4}	0.18%	8.0×10^{-6}
urine	X - 16563	urine	X - 17739	0.02	3.5×10^{-1}		−0.03	1.8×10^{-5}	0.83%	8.0×10^{-6}
plasma	S-methylcysteine	urine	S-(3-hydroxypropyl)mercapturic acid (HPMA)	0.15	1.3×10^{-15}	0.06%	−0.02	3.3×10^{-2}		8.2×10^{-6}

Figure 8.4 Differential metabolic multi-fluid network. Using permutation testing, I identified 86 edges that differ significantly between the metabolic models of cases and controls. These fall into three categories: edges that only are present in the control network (13 edges, yellow), edges that only are present in the case network (55 edges, red) and edges that are present in both networks but with different strengths (18 edges, black).



8.3 Discussion

I analysed the associations of advanced kidney disease with 882 metabolites across three fluids in a cohort of 978 individuals. This is the first comprehensive analysis combining metabolic changes across three fluids and the first multi-variate analysis of metabolic processes associated with renal function. This includes the first holistic study on salivary metabolite changes with respect to renal function and it extends earlier analyses of plasma and urine metabolomes by combining results across fluids in an integrated statistical approach.

Metabolic differences between CKD patients and controls were widespread across fluids and pathways. Sekula et al. (2016) previously analysed the association of renal function with blood metabolites in the TwinsUK and KORA studies and found 54 metabolites significantly associated with eGFR after meta-analysis. 42 of those were included in my analysis and 36 of them were significantly associated with CKD. While the associations were largely similar between these two studies (Figure E.1), there were several exceptions. Interestingly, urea and urate are associated with renal function in the general population and O-sulfo-L-tyrosine even predicted incident CKD (Sekula et al., 2016) but neither of them differed between cases and controls here. This might be due to successful filtration of those metabolites by dialysis or successful renal replacement therapies. In contrast, 74 plasma metabolites were significantly associated with CKD only here, including 1,5-Anhydroglucitol (1,5-AG) and hypoxanthine. 1,5-AG is a marker of glycaemic control (Yamanouchi et al., 1994) and – more interestingly – a marker for tubular reabsorption (Saito et al., 1996), which is impaired in advanced stages of CKD, as analysed here.

While these associations are of potential interest, the majority of analysed metabolites differed between cases and controls and the univariate tests do not provide the means to identify the underlying metabolic processes that lead to these widespread changes. In contrast, GGMs infer models of the metabolism, rather than considering changes of individual metabolite levels. The resulting models are sparse with only 2.8 % and 1.6 % of edges being included in controls and case networks, respectively. Differential network analysis revealed 86 edges of interest that differ significantly between the networks. It is striking how sparse the metabolic GGMs and particularly the differential network are, with only 0.022 % of all edges (1.4 % of the tested edges) being significantly different. This indicates that the underlying, independent metabolic disruptions are rather small compared to the overall effect on the metabolism. Most differential edges fell into three categories: (i) salivary amino acids, (ii) steroid hormones, and (iii) xanthine metabolites, which I will discuss in the following.

8.3.1 Salivary branched-chain amino acids

I found salivary BCAA metabolism among the most affected pathways (Figure 8.4c). Leucine and isoleucine in saliva are the nodes with most differential edges (4 each), followed by salivary leucine (3) and valine (2).

BCAA metabolism has been long known to be affected by renal function (c.f. chapter 6), partially due to the paramount importance of renal reabsorption and excretion of amino acids (Tizianello et al., 1983; Verrey et al., 2009). Also, nutritional supplementation with BCAAs has

been suggested to improve appetite (Cano et al., 2006) and the outcome (Bolasco et al., 2011) in dialysis patients. Accordingly, valine and leucine (but not isoleucine) were significantly depleted in plasma of patients ($p = 1.2 \times 10^{-12}$ and $p = 1.4 \times 10^{-7}$, respectively). However, on the one hand their concentration in saliva did not significantly differ between the groups ($p > 0.03$, $p > 0.19$), and on the other hand there were no differential edges in the GGM between amino acids in blood. Thus, neither nutritional supplements, which would increase amino acid concentration in saliva, nor decreased salivary excretion due decreased levels in blood are plausible causes for the observed differential correlations.

One potential cause of the observed changes in salivary BCAA metabolism are changes of the oral microbiome composition. Evidence suggests that the oral microbiome of CKD patients differs from healthy individuals (Araújo et al., 2015). With microbes producing and metabolising many amino acids and particularly BCAAs (Takahashi, 2015), these changes could explain the observed changes in salivary amino acid metabolism. Also, I found a differential edge linking the salivary short chain fatty acids (SCFAs) caproate and valerate ($p = 2.0 \times 10^{-7}$), which are also common microbial metabolites. This illustrates the potential of my approach to highlight disruptions of metabolic process, independently of changes of metabolite concentrations. However, without microbial sequencing data available, these changes have to be confirmed by future studies. Nevertheless, my findings might support future studies on the comorbidity of renal disease and periodontal disease (Ariyamuthu et al., 2013).

8.3.2 Steroid metabolism

I found several links of the steroid metabolism significantly different between cases and controls. For instance, the connection of cortisol in plasma and urine ($p = 4.0 \times 10^{-7}$) was observed in controls only (Figure 8.4b). Similarly, pregn steroid monosulfate in plasma was connected with pregnen-diol disulfate and 21-hydroxypregnenolone disulfate in urine, respectively, only in controls. In contrast, the edge between pregn steroid monosulfate and andro steroid monosulfate was only present in the network of cases but not the control graph ($p = 8.0 \times 10^{-6}$, Figure 8.4a). Cortisone and cortisol also showed differential edges with several unknown metabolites, including X-15636, X-17339 and X-12776, which might be identified in the future.

Glucocorticoids, such as cortisol, as well as sex hormones, such as androsterone, are produced by the suprarenal glands (Arlt et al., 2005). Glucocorticoids generally have a catabolic effect on the metabolism (Lorraine et al., 2003), mainly leading to increased availability of energy substrates, e.g. by increasing hepatic glucose (Lecocq et al., 1964) and fatty acid (Peckett et al., 2011) output. They also increase urinary flow and glomerular filtration (Lorraine et al., 2003) in the kidneys. However, corticosteroids have multiple and diverse effects. For instance, they act on the hydro-electrolyte equilibrium, affecting the reabsorption of sodium and water in the tubular lumen (Hunter et al., 2014) and play an important role as immunomodulators and anti-inflammatory agents (Coutinho et al., 2011). Hence, changes of steroid metabolism have the potential to cause the severe metabolic anomalies of ESRD. Downstream changes, for instance in electrolyte concentrations, are a potential cause for comorbidities of renal disease, such as hypertension (Krakoff, 1988) and cardiovascular disease (Walker, 2007).

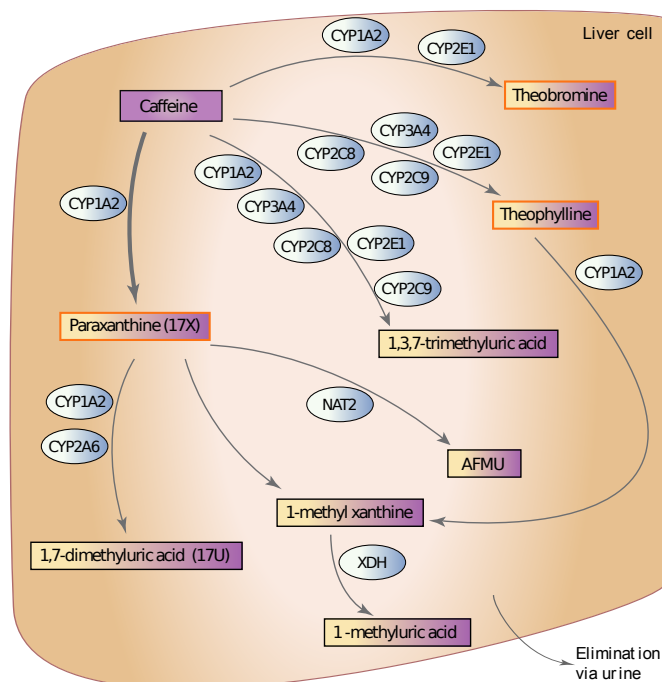
However, the observed changes in steroid hormone metabolism could also result from drugs given to patients. Anabolic steroids have been shown to facilitate weight gain and increase muscle mass and power in dialysis patients (Johansen et al., 1999) and improve overall survival (Boero et al., 2000). Also, steroids are used for immune suppression, for instance for IgA nephropathy (Lv et al., 2012), and steroid treatment might be continued despite dialysis (Altieri et al., 2002).

8.3.3 Xanthine metabolism

Lastly, I found significant changes of the xanthine metabolism. The link between plasma and urine 7-methylxanthine was stronger for cases than controls ($p < 1.0 \times 10^{-7}$). In contrast, the links between plasma paraxanthine and urine 1-methylxanthine ($p = 2.0 \times 10^{-7}$) as well as between urinary 7-methylxanthine and 1-methylurate ($p = 2.0 \times 10^{-7}$) were significantly stronger in controls. Also, the link between caffeine in plasma and saliva was significantly different between the groups ($p = 2.0 \times 10^{-7}$).

Methylxanthines are caffeine metabolites (Callahan et al., 1982), which are produced by several *cytochrome P450* (CYP) enzymes, particularly CYP1A2, and NAT2 (Figure 8.5) (Thorn et al., 2012). Differential GGM edges between plasma and urine concentrations of xanthines suggest changes in these degradation processes, leading to altered catabolism and excretion of methylxanthines in CKD patients. While the metabolism of coffee does not seem to be the most important effect of renal disease in the first place, there are several interesting aspects to it:

Figure 8.5 Caffeine Metabolism. Caffeine is almost completely metabolised, predominantly by demethylation to paraxanthine or theobromine. The end products are excreted through the kidneys in urine. Squares indicate metabolites and ellipses represent genes/proteins. (Picture reproduced with permission of PharmGKB and Stanford University).



First, altered caffeine metabolism is a marker for generally impaired metabolism of drugs. Decline of kidney function clearly leads to impaired renal detoxification of drugs and xenobiotics, thus altering their bioavailability and effect on the organism (Doogue et al., 2011; Dreisbach et al., 2008). Moreover, CKD also affects hepatic clearance of drugs, most notably the CYP metabolism (Leblond et al., 2000; Dreisbach et al., 2008), by reducing enzyme expression levels (Leblond et al., 2001). This might cause the changes in coffee metabolism observed in this study. However, caffeine likely is just a marker of these molecular changes that is easy to detect due to its high prevalence. Other, less common and lower abundant drugs might be affected in the same way, though are harder to detect on a broad metabolomics platform.

Secondly, there is ample evidence for a detrimental effect of methylxanthines on renal function. Xanthines can increase urine flow in a dose-dependent manner (Osswald et al., 2011) and increase excretion of electrolytes, lithium (Shirley et al., 2002), and cyclic adenosine monophosphate (cAMP) (Coulson et al., 1989). This is probably a consequence of the inhibitory effect of methylxanthines on the adenosine receptor, which causes a reduction in renal reabsorption (Rieg et al., 2005). Inhibition of adenosine receptors is also thought to be the cause for methylxanthine-induced increase of renin (Balakrishnan et al., 1993), and consequently the activation of the renin-angiotensin-aldosterone system, which in turn increases hypertension and sodium reabsorption (Choi et al., 1993; Ohnishi et al., 1986). While this is only a superficial and incomplete list of the effects of methylxanthines (which have been thoroughly reviewed before, e.g. by Osswald et al. (2011)), it illustrates that changes of xanthine metabolism are actually more than a mere consequence of CKD. Xanthines also are uremic toxins that have the potential to cause the widespread metabolic shift as well as comorbidities of renal function.

8.4 Conclusion

My results support previous studies that observed the enormous metabolic changes observed in CKD patients across different fluids. Univariate testing identified significant differences of 621 of the 882 analysed metabolite levels. However, metabolites interact and disrupting one metabolic process will change the kinetics of adjacent reactions in the metabolic network, which will in turn disrupt their adjacent processes and so on. Univariate tests help to identify the complete set of affected metabolites, however they fail to identify the underlying disruption.

In contrast, differential metabolic network analysis takes into account the hierarchical structure of the metabolic network. Hence, GGMs can identify metabolic anomalies that are independent of all other metabolites and potentially account for the widespread changes. Here, differential metabolic network analysis emphasised three pathways that differ between cases and controls: branched chain amino acid metabolism in saliva, steroid hormone metabolism and excretion in urine, and xanthine metabolism. Particularly the two latter have the potential to cause widespread metabolic changes.

Despite the promising results, this study has several limitations. First, a relatively small number of cases was available for analysis. More importantly, the patients are rather diverse with half of them receiving dialysis and all of them receiving various treatments. Thus, I cannot differentiate disease effects from treatment effects. Also, causality cannot be inferred from the results, as data is observational only and cannot distinguish causes and consequences of the disease. Finally, results presented here should be considered preliminary and further analyses, for example comparing different treatment groups, particularly dialysis and steroid treatments, and genetic predisposition, are yet to be conducted. Also, I excluded 782 metabolites due to their high proportion of missing values, though 208 of them are significantly different between cases and controls (Figure E.4). Analysing accumulation of rare metabolites in CKD patients might shed light on mechanisms causing comorbidities of renal disease. However, many (75.5 %) of these rare metabolites are unknown and difficult to interpret.

Nevertheless, this approach demonstrates the potential of systems biology to integrate data across fluids, facilitating conclusions on inter-fluid transport processes, and to highlight the relevant pathways, that can subsequently be followed up in subsequent experiments. As such follow-up studies are very much limited by the effort required to perform in vitro experiments, mouse experiments, or randomised trials, reducing the number of 'interesting' targets is crucial. Here, I identified three pathways that are significantly altered in CKD patients. Particularly steroid and xanthine metabolism are – based on prior knowledge – potential driving factors for metabolic disruptions as well as development of comorbidities and, thus, promising targets for further experiments and potential drug targets.

CHAPTER 9

Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model

Ageing has widespread effects on the organism, thus increasing susceptibility to diseases. Here, I analysed the underlying molecular changes and how they mediate disease comorbidities by integrating ageing markers from four different 'omics' datasets with extensive phenotypic data. I inferred a graphical model to investigate the multi-variate dependencies of molecular and clinical phenotypes.

Part of this work has been published in *Scientific Reports* (Zierer et al., 2016b), the published version is attached in appendix G.4.

Ageing is a multi-factorial process that affects organisms at multiple levels. In previous chapters, I investigated metabolic and glycomic profiles associated with biological ageing and specifically CKD. However, these molecular changes do not occur separately but they influence each other, as both ageing and CKD are multi-factorial processes that affect the entire organism at multiple levels. Taking into account the complex interplay between the different levels is necessary to unveil the causal structure of multi-factorial processes.

Various concepts have been proposed to integrate data from different omics technologies, thus shaping the newly emerging fields of systems biology, systems genetics (Civelek et al., 2014), and systems medicine (Gustafsson et al., 2014). Thereby, networks have been shown to be particularly useful to assess complex interactions in a dataset and to illustrate multivariate dependencies (see section 1.3.1). Graphical models (GMs) facilitate the inference of networks from measured data and investigate multivariate dependencies of the included variables. Although the direction of these associations and hence causality can in most cases not be determined by these models, the resulting network of direct associations between variables can be considered as the undirected skeleton of their underlying causal structure.

In this chapter, I investigated the molecular basis of age-related diseases, aiming to identify molecular mediators that lead to disease comorbidities. To this end, I used an integrated mixed graphical model (MGM) approach to combine ageing markers from four different high-throughput omics datasets on the same individuals, namely epigenomics, transcriptomics, glycomics, and metabolomics, together with extensive phenotypic data. The resulting network of direct associations that are independent of all other variables within the model is expected

to provide valuable insights into the direct molecular interdependencies between various age-related phenotypes, despite the lack of causal inference. To the best of my knowledge, this is the first study using graphical models to combine data from multiple molecular omics and phenomics datasets.

9.1 Methods

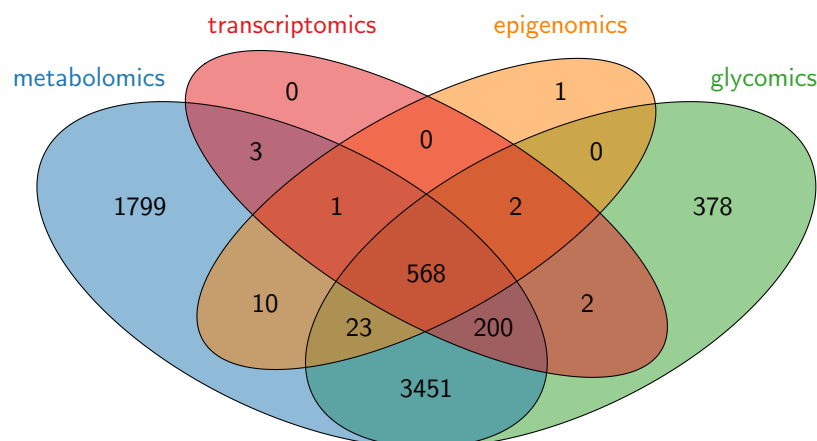
9.1.1 Study Population

For this study, I analysed data of 510 female participants of the TwinsUK cohort (section 3.1.1) – 62 MZ twin pairs, 116 DZ twin pairs and 154 singletons – aged between 34 and 84 (mean 59.0 ± 9.4) with measurements for epigenomics, transcriptomics, glycomics, and metabolomics as well as extensive clinical information available (Figure 9.1).

9.1.2 Data Acquisition and Processing

I combined four different high-throughput ‘omics’ datasets: epigenomics, transcriptomics, glycomics and metabolomics. Network inference is not feasible for the combined dataset of several hundred measured metabolites, thousands of RNA transcripts and hundreds of thousands of CpG sites, particularly given the limited number of samples. Therefore, I used a knowledge-driven approach to reduce the number of variables from each dataset. To this end, I selected only variables that were previously reported to be strongly and independently associated with chronological age as described in the following (and listed in Table F.1).

Figure 9.1 ‘Omics’ datasets in the TwinsUK cohort. Several high-throughput omics datasets are available for the TwinsUK cohort (section 3.1.1 for details), however all of them were measured in different subsets of the full cohort. I selected 568 individuals with four different ‘omics’ measurements available. 510 of them, with comprehensive data on various phenotypes available, were used to infer the Age-MGM.



9.1.2.1 Epigenomics

DNA methylation levels were measured in adipose tissue samples using the Infinium Human-Methylation450 BeadChip (Illumina Inc., San Diego, CA) as previously described (Grundberg et al., 2013). Data was corrected for technical variation using the BMIQ method and corrected for batch effects and bisulfite conversion levels using linear mixed effect models.

Weidner et al. (2014) showed that only three ageing-related differentially methylated regions (aDMRs) are enough to predict the chronological age with high precision. Those three sites, namely cg02228185 (in *aspartoacylase* (ASP)), cg25809905 (in *integrin subunit alpha 2b* (ITGA2B)), and cg17861230 (in *phosphodiesterase 4C* (PDE4C)), were selected for further analyses.

9.1.2.2 Transcriptomics

RNA abundance was measured in abdominal fat samples using the Illumina Human HT-12 V3 Bead chip (Illumina Inc., San Diego, CA) as part of the MuTHER project as previously described (Grundberg et al., 2012). The probe intensities were adjusted for batch effects using linear models prior to analysis.

A previous study found 188 genes (199 probes) significantly associated with chronological age (Glass et al., 2013). I performed stepwise regression to select expression probes independently associated with age. This procedure left 24 probes from 24 different genes (Table F.1) for further analysis.

9.1.2.3 Glycomics

IgG glycosylation was measured by Genos Glycoscience and normalised as described in section 3.3.2.

It has been shown that a linear combination of only three IgG glycan structures – GP6, GP14, and GP15 – explains 58 % of the variance in age (Kristic et al., 2014) and furthermore correlates with several age-related associated phenotypes. These three structures were selected for my network analysis.

9.1.2.4 Metabolomics

For this project, I used metabolomics measurements from the V3 version of Metabolon's untargeted platform as described in section 3.3.1.

About half of all known circulating blood metabolites were reported to be associated with chronological age in several large population studies (see section 1.2.6). I selected 22 of these metabolites, which were found to be independently associated with age and together explain 59 % of the variance of chronological age (Menni et al., 2013b).

9.1.2.5 Clinical Phenotypes

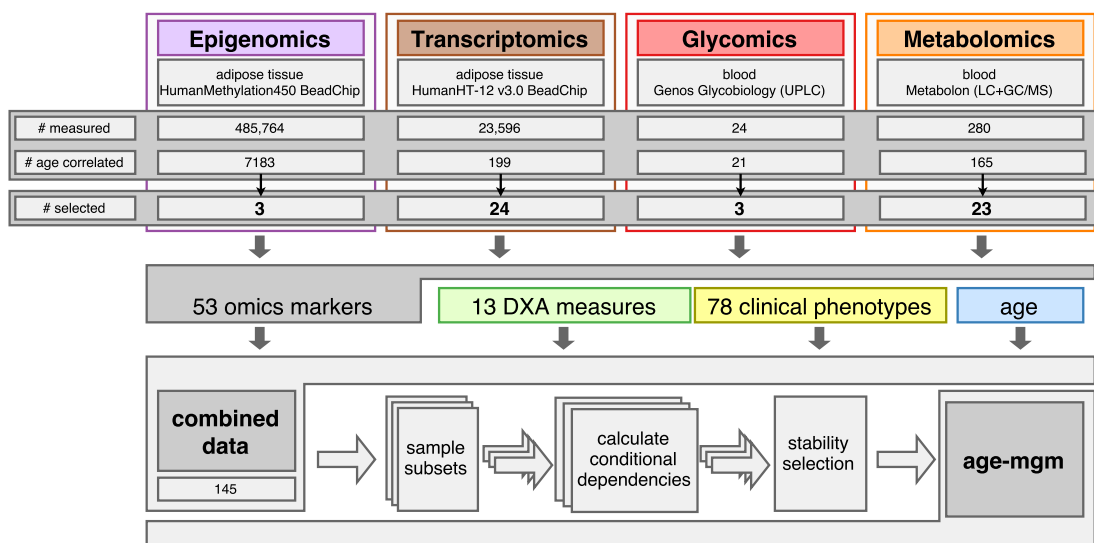
A total of 92 phenotypes were combined with the previously described omics data (Table F.2). Besides chronological age, I included anthropometric measures, such as height, weight, and BMI (see section 3.2.1). Additionally, 13 body composition variables, measured by DXA scans (see section 3.2.2), were used to comprehensively assess body composition. The eGFR, calculated from serum creatinine levels, was used as measure of renal function (section 3.2.3); circulating levels of GGT and ALT were used as markers of liver function (section 3.2.5). Lung function was measured as FEV1 and FVC (see section 3.2.6).

Moreover, I included data from various questionnaires assessing disease states, such as arthritis, asthma, and chronic pain. Questionnaires were also used to collect lifestyle parameters, such as smoking and physical activity. Information on food intake was collected using an established food frequency questionnaire (FFQ) (Bingham et al., 2001). Item frequencies were merged into 54 food groups and transformed into orthogonal patterns using PCA (Teucher et al., 2007). In the ageing model, I used the first five principal components, which correspond to five different dietary patterns (Supplementary Table F.3).

9.1.2.6 Data Pre-processing.

I excluded samples with more than 20 % missing values and subsequently excluded variables with more than 20 % missing values, leaving 510 individuals in the final dataset. The remaining missing values in both omics variables and clinical phenotypes were imputed using the mice package (Buuren et al., 2011) (see section 3.4.5). All continuous variables were inverse normalized and categorical variables were dichotomized. To account for family relatedness, I included the unique family identifier as additional variable during the network inference and removed the according node from the network prior to analysis.

Figure 9.2 Age-MGM variable selection and inference process. The flowchart illustrates the selection of variables from the four different omics datasets and the inference process.



9.1.3 Network Inference

The mixed graphical model (MGM) was inferred using the graphical random forest (GRaFo) method (Fellinghauer et al., 2013) with the complementary pairs stability selection (CPSS) modification (Shah et al., 2013) as described in section 3.4.8.

As effect estimators of random forest, and partial effects in MGMs in general, are non-linear and context-dependent, there is no estimator for the sign of an edge in my model. I, thus, inferred the signs from regression models, regressing each variable against all others, for visualization purposes.

9.1.4 Network Analysis

I analysed the resulting MGM as undirected, unweighted network. Several measures, including node degree, clustering coefficient, and betweenness centrality, were calculated to describe topological features of the network as whole and individual nodes, as described in section 3.4.9. All analyses were performed in R using the *igraph* package.

9.1.5 Network Stability

To test the robustness of the model I investigated the dependence of the network topology on the inference process.

9.1.5.1 Edge inclusion cut-off

Firstly, I assessed the robustness of node centrality and module assignments when varying the cut-off for edge inclusion. To this end, I defined different models by including edges that were contained in 20 %, 40 %, 60 %, 80 %, and 100 % of the subsamples, respectively, where 80 % corresponds to the original model. Additionally, I analysed a weighted network (Zhang et al., 2005a) including all edges that were observed in at least one sample.

As a measure of stability of node centrality, I calculated the correlation of node degrees and clustering coefficients between the original model and the model in the networks for different edge cut-offs (Figure F.3).

To assess the stability of module assignments, I calculated the adjusted RAND index (Rand, 1971; Hubert et al., 1985) as a measure of similarity between the seven network modules of the original Age-MGM with modules identified from the networks that were inferred based on different edge cut-offs. The RAND index quantifies the similarity of module assignments by counting the agreements between two different module assignments and adjusting it for the number of agreements that are expected by chance. An adjusted RAND index of 1.0 indicates identity between the module assignments of two networks while values close to 0.0 indicate dissimilarity of the assigned modules. In addition, I compared the adjusted RAND indices of the networks for the different edge cut-offs with the background distribution of 1000 randomly sampled module assignments (Figure F.3c).

9.1.5.2 Variable selection

Secondly, I investigated the stability of the network and particularly the module assignments depending on the pre-selection of omics variables prior to the model inference. To assess the influence of this selection step on the final results, I inferred a second model from the same dataset but including all metabolomics variables with known chemical identity, thus, completely dispensing variable selection for the metabolomics data.

To compare module assignments for the large network with the assignments for the original Age-MGM, I restricted the large network to the nodes of the Age-MGM. Edges in this network represent conditional dependence, given all other variable in the Age-MGM and given the 196 additional metabolites. Module assignments were then compared using the adjusted RAND index. A detailed comparison module memberships is presented in Table F.4.

9.1.5.3 Stability across datasets

Finally, I assessed the stability of the Age-MGM with respect to the generating dataset by comparing it with two models inferred from two disjoint datasets containing either the first or the second twin of each family, respectively. Singletons were distributed randomly across both datasets (Figure F.4).

9.2 Results

I inferred a MGM using observational data from a cohort of 510 women, aged between 34 and 84, integrating selected age-associated markers from four different ‘omics’ datasets with 92 clinically assessed phenotypes (Figure 9.2).

The final model consisted of 145 nodes and 316 undirected edges connecting them (Figure 9.3). Thus, it was much sparser (316 edges instead of 1900) than a regular correlation graph based on significant pairwise correlations of variables from the same dataset (Figure F.1). Most of the nodes (96) formed one large connected component, which I refer to as Age-MGM. There were two smaller components of 8 and 4 nodes of variables related to pain and memory function, respectively, two isolated pairs of nodes and 33 unconnected nodes. The degree, betweenness, and clustering coefficients of all nodes in the network are presented in Tables F.1 and F.2.

9.2.1 Topological properties of Age-MGM

The large connected component Age-MGM contains 96 variables including age, along with variables from all four ‘omics’ datasets, and 286 edges connecting them. It has an average node degree of 6.0, an average local clustering coefficient of 46.6 %, and an average shortest path length of 3.2.

Corresponding to the high local clustering and the short average path lengths in the model, its small-world index is 6.1, indicating a marked small-world-ness. Removing age from the network does not reduce the small-world-ness of the network. In comparison, the correlation graph, restricted to the same vertices as in the Age-MGM (Figure F.1), has a just slightly higher clustering coefficient of 57.0 % despite the much higher average node degree of 31.2, which results in a small world-index of only 1.7.

As expected, age is the most densely connected node with a degree of 27 (Figure 9.4a). It has a low clustering coefficient (8.0 %) but high betweenness centrality (47.5 %), indicating that age connects different clusters, while its neighbours tend to be unconnected. With an average shortest path length of 2.1 age is also the most central node in the Age-MGM.

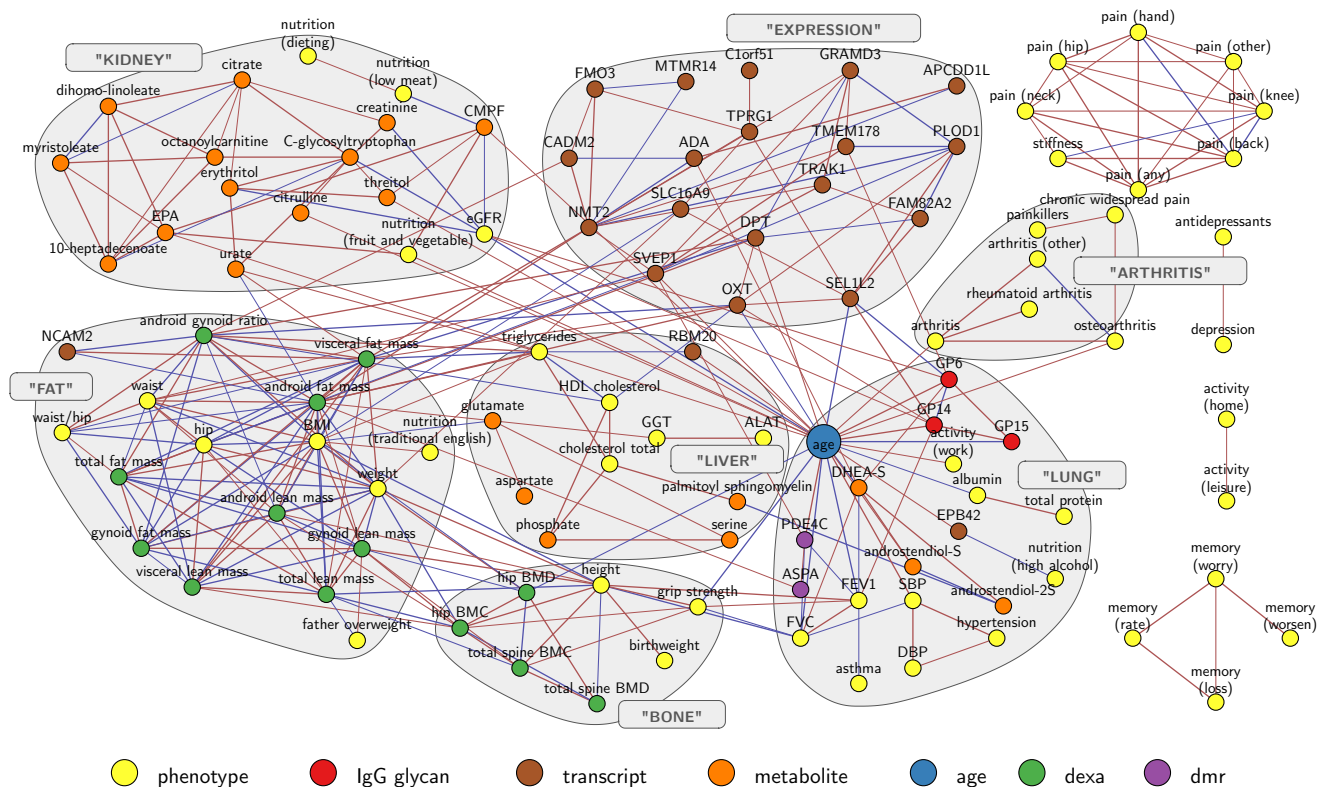
9.2.2 Modularity of the Age-MGM

There are more edges between variables originating from the same ‘omics’ dataset than edges connecting them. Particularly transcriptomics and metabolomics variables form dense clusters with 37 and 34 edges within them, respectively. In contrast, only 7 edges connect transcriptomics and metabolomics variables with variables from other omics sets. Similarly, the body composition variables measured by DXA are densely connected by 45 edges amongst them (Figure F.2).

To analyse the graph structure in an unbiased way, independently of the different experimental methods, I used a modularity-based algorithm for cluster detection (see section 3.4.9). This approach yielded seven modules (Figure 9.3).

The first cluster (*EXPRESSION*) contains all but three gene expression markers. It is connected with neighbouring clusters mainly via expression levels of *oxytocin/neurophysin I prepropeptide* (OXT), which has 6 edges outside of its cluster (Figure 9.4b), and *sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1* (SVEP1), which has the highest betweenness centrality (10.5 %) within the cluster. The second cluster (*LUNG*) contains age and several of its direct neighbours from different ‘omics’ layers. The lung function parameters FEV1 and FVC were the most densely connected phenotypes in the cluster (degrees 8 and 7, respectively). Both were embedded in a tight cluster with local clustering coefficients of 35.7 % and 47.6 %, respectively (Figure 9.4c). Age is also connected to another small cluster of arthritis-related variables (*ARTHRITIS*). The body composition variables fall in two different clusters, one of them containing bone-related variables (*BONE*) and the other fat and lean mass-related variables (*FAT*). While the *BONE* cluster is densely connected with the *LUNG* cluster, all connections between the *FAT* cluster and the *LUNG* cluster, which contains age, were mediated, mainly via the *EXPRESSION* cluster. The next cluster (*LIVER*) contains the liver markers ALT and GGT along with cholesterol and triglyceride levels and several amino acids. It also contains the gene expression marker of the *RNA binding motif protein 20* (RBM20) gene, which mediates the connection of the cluster with age and the *LUNG* cluster. The last cluster (*KIDNEY*) contains mainly metabolite levels, but also markers of nutrition and the eGFR. With 9 edges, C-glycosyltryptophan is central within the metabolite cluster. However, the eGFR (degree 7) is the main connection of the metabolomics cluster with age as well as IgG

Figure 9.3 Ageing graphical model. Each node in the graph represents one age-related variable. Omics markers were selected according to literature from epigenomics (purple), transcriptomics (brown), glycomics (red), and metabolomics (orange) datasets and combined with DXA measurements (green) and other clinical phenotypes (yellow). Edges between nodes were inferred using a mixed graphical model approach, and thus indicate the conditional dependence between variables; the colour represents positive (red) and negative (blue) correlation. An unbiased cluster detection algorithm was used to identify densely connected modules within the network, indicated by grey borders.

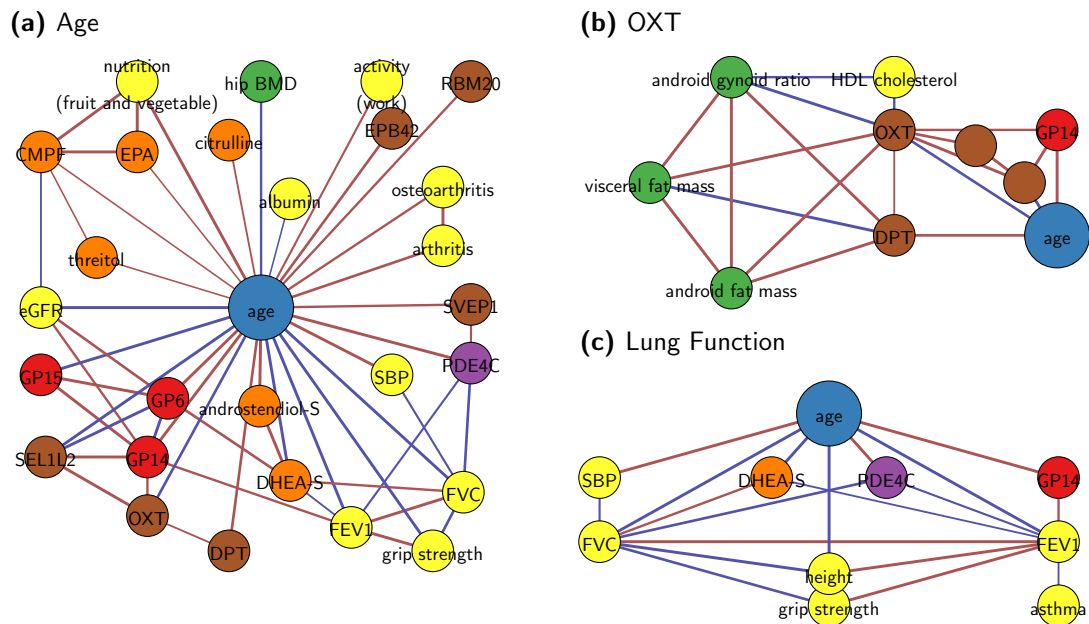


glycosylation markers. The only connections of the *KIDNEY* cluster with other clusters, other than with age, were edges between the metabolite eicosapentaenoic acid (EPA) with triglyceride levels and urate with the *FAT* and *LIVER* clusters

9.2.3 Robustness of the Age-MGM

To assess the robustness of the model, I inferred additional networks based on different cut-offs for edge inclusion. Comparing these networks with the Age-MGM, I found node centrality as well as module assignments very stable against variation of the inference process (Figure F.3). Also, the modules of the original Age-MGM remained stable even when including all available metabolomics variables with known chemical identity, i.e. more than doubling the number of variables in the model (Table F.4). The resulting graph consisted of 341 nodes (145 of them from the original model and 196 newly added metabolites) connected by 1152 edges. 707 of these edges were amongst the new metabolites, 174 connected one new metabolite

Figure 9.4 Ageing network modules. Each panel shows one sub-graph of the Age-MGM (Figure 9.3). (a) The direct neighbourhood of chronological age. (b) The hormone *oxytocin/neurophysin I prepropeptide* (OXT) mediates association of fat mass variables with age as well as the *immunoglobulin G* (IgG) glycosylation marker glycan peak 14 (GP14). (c) The direct neighbourhood of the lung function measures forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) contains three ‘omics’ markers: dehydroepiandrosterone sulfate (DHEA-S), *phosphodiesterase 4C* (PDE4C) and the GP14, which represents the proportion of FA2G2 glycan in total IgG glycans.



with one original variable and 271 edges were amongst original variables, of which 253 were also in the original model. The 63 edges that were missing in the large network compared to the original Age-MGM were, on average, contained in 58 % of the subsamples of the large network, suggesting that they were excluded due to the limited power. The added metabolites were predominantly peripheral in the Age-MGM, with 160 of the 174 edges connecting new metabolites with original variables linking them with metabolites and the remaining 14 with either blood lipid measures or renal function.

To investigate the reproducibility of the Age-MGM in different datasets, I calculated two separate models from disjoint datasets incorporating only the first and second twin of each family, respectively. The two resulting models reproduced 93.5 % of all edges, with only 21 edges being unique to the original model (Figure F.4). These 21 unique edges were on average included in 78 % of the subsamples of two individual twin models, and thus, just missed the edge-inclusion cut-off of 80 %, potentially due to lower power in the smaller datasets.

9.3 Discussion

In this chapter, I inferred a robust graphical multi-omics model of age-related diseases by integrating disease phenotypes with molecular markers from four ‘omics’ layers based on data of 510 women from the TwinsUK cohort. This enabled conclusions on molecular changes that mediate disease comorbidities.

Despite the sparsity of the model, which omits mediated associations, most variables formed one connected component, the Age-MGM, consisting of seven modules. Each of these modules represents a different aspect of ageing, such as metabolic ageing linked to decline of renal function (*KIDNEY* cluster), the change of fat and lean mass (*FAT* cluster) – along with the closely related changes of gene expression in adipose tissue (*EXPRESSION* cluster) – and the decrease of BMD and bone mineral content (BMC) (*BONE* cluster) (Figure 9.3).

The Age-MGM models multivariate dependencies of age-related diseases that potentially underlie the commonly observed comorbidities. Edges in this model represent conditional dependence between two variables, while the absence of an edge implies their conditional independence given all other variables in the model. Specifically, this means that previously observed age-associations of the variables, which are not directly linked to age in this model, occur due to the mediation by other factors. This differentiation between mediated and direct associations permits conclusions on underlying mechanisms even though the causal directions cannot be inferred.

9.3.1 Decline of renal function links age with metabolic shift

The blood metabolome was shown to be strongly influenced by age in several studies (section 1.2.6). In the Age-MGM, most of the age-associated metabolites (13) formed one large cluster with only four of them being directly linked to age, while the remaining nine metabolites were only indirectly associated with age. For six of these nine metabolites, the shortest path to age was mediated by the eGFR, a measure of renal function. Even though the model is undirected, age was the only non-modifiable variable in the model. I, thus, hypothesize that with increasing age renal function declines leading to the major shift in the ageing blood metabolome, which possibly causes further diseases.

9.3.2 Urate mediates association of renal function with body composition

Urate mediated the connection of the *KIDNEY* cluster with *FAT* and *LIVER* clusters. Hyperuricemia has been previously reported to be associated with obesity, particularly increased visceral fat mass (Takahashi et al., 1997) and increased triglyceride levels (Giacomello et al., 1997), which I find directly linked in the Age-MGM. Indeed, there is evidence that urate actually contributes to the development of obesity and diabetes, rather than being just a consequence: Elevated serum levels of urate were found to predict obesity (Masuo et al., 2003) and diabetes (Nakanishi et al., 2003). Moreover, by knocking out the uric acid transporter *solute carrier family 2, facilitated glucose transporter member 9* (SLC2A9) in mice, DeBosch et al. (2014) and colleagues found that hyperuricemia causes several phenotypes of the metabolic syndrome,

including obesity, dyslipidaemia and hypertension. Administering a compensating treatment attenuated some but not all the observed symptoms.

Hyperuricemia is also a known comorbidity of renal disease, however the causal direction of this association is controversial (Johnson et al., 2013). Moreover, renal disease and hyperuricemia were also shown to affect the gut microbiome composition (Vaziri et al., 2013), which is known to be strongly associated with obesity and other symptoms of the metabolic syndrome (Parekh et al., 2015) (see chapter 7). Thus, the microbiome is a potential hidden mediating factor of the association between hyperuricemia and obesity.

Even though its mode of action remains elusive, my model suggests urate to be a key factor for the comorbidity of renal disease with symptoms of the metabolic syndrome and obesity.

9.3.3 Lung function as a central ageing process

Lung function is a central process in the Age-MGM. Both lung function measures FEV1 and FVC are directly connected with age and are, besides age, the most densely connected nodes in their cluster. They are connected with three different 'omics' markers (Figure 9.4c): First, the metabolite dehydroepiandrosterone sulfate (DHEA-S) is one of the most abundant hormones in humans that is well known to decrease with age (Orentreich et al., 1992) and even has been suggested as an anti-ageing drug (Baulieu et al., 2000). Moreover, DHEA-S has been found to prevent and even revert pulmonary hypertension in rats (Bonnet et al., 2003), suggesting a causal effect of DHEA-S on lung function.

Second, the methylation probe cg17861230 lies in the PDE4C gene, an enzyme that catalyses the hydrolysis of cAMP. Expression levels of PDE4C were previously found to be associated with lung function (Tang et al., 2006). PDE4 is a target for drugs against chronic obstructive pulmonary disease (COPD) and one PDE4 inhibitor, Roflumilast, has already been approved by the EMA for treatment of COPD (Calverley et al., 2009). This highlights the potential of my model to emphasise causal links, by removing spurious correlations.

Third, the IgG glycosylation marker GP14, which represents the percentage of FA2G2 glycan in total IgG glycans, is connected to lung function in the Age-MGM. GP14 is a glycan structure with terminal galactose, which is known to change the inflammatory state of IgG (Karsten et al., 2012). While defects of general protein glycosylation (Nihlén et al., 2001) as well as an involvement of IgG (O'Keeffe et al., 1991) in COPD have been previously reported, glycosylation of IgG has so far not been associated with lung function.

My model identifies two known causes of lung disease, the hormone DHEA-S and the gene PDE4. It moreover suggests a contribution of IgG mediated inflammation in the age-related decline of lung function. As IgG glycosylation is also related with kidney function in the Age-MGM as well as in previous studies (see chapter 5), this provides a potential explanation for the comorbidity of lung disease and renal disease.

9.3.4 Hormone expression directly associated with body composition

It is commonly known that BMI as well as waist and hip circumferences and body fat mass change with age. Nonetheless, I found neither of them directly linked to age in the Age-MGM. Instead, all associations between age and the *FAT* cluster are mediated. One of these mediation paths leads via urate and renal function (as discussed above). A second path leads via the *EXPRESSION* cluster and, particularly, the expression of OXT (Figure 9.4b), which accordingly mediates 6.0 % of all shortest paths in the model. OXT is also directly linked to HDL cholesterol levels.

While adipose tissue was traditionally considered as storage tissue, it receives increasing attention as endocrine organ (Hauner, 2004), which amongst others produces OXT. OXT is a hormone with a broad spectrum of functions, ranging from reproductive functions and control of social behaviour (Bartz et al., 2011) to energy metabolism (Chaves et al., 2013). One common explanation for the influence of OXT on obesity is its effect on food intake (Lawson et al., 2015), though there also is a diet-independent effect of OXT on the lipid metabolism (Deblon et al., 2011). Thus, OXT was suggested as drug against obesity and T2D development and has been successfully tested in a first pilot trial (Zhang et al., 2013).

My results indicate that the age-related change of body composition can partially be attributed to alterations of gene expression in adipose tissue and particularly to a change in OXT expression, independently of food intake. OXT might also drive common comorbidities of obesity by causing dyslipidaemia, which in turn increases the risk of cardiovascular diseases (Arca et al., 2007).

9.3.5 IgG glycosylation as new mechanism of obesity-associated inflammation

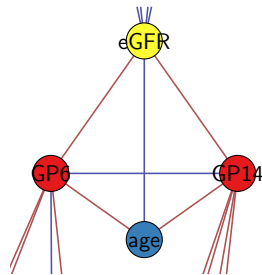
Obesity is known to be associated with chronic low-grade inflammation and activation of immune function (Fogarty et al., 2008), which is thought to be an important mediator between obesity and its common complications, such as T2D (Esser et al., 2014).

In my model the expression of OXT mediates the association of android and visceral fat mass with inflammatory IgG glycosylation. The influence of OXT on IgG might be mediated by *interleukin-6* (IL6), which was found to be down-regulated by OXT in vitro (Szeto et al., 2008) and thus causes decreased IgG production in B-cells (Maeda et al., 2010). My model confirms an effect of increased fat mass on IgG, mediated by OXT, in vivo. Moreover, it provides evidence that OXT also affects IgG glycosylation in addition to its expression, thus altering its inflammatory potential. I hypothesize that this is a new mechanism of obesity-induced inflammation that appears to be independent from previously identified pathways that are mediated by leptin or adiponectin (Tilg et al., 2006). Both are co-expressed with OXT in this data (Pearson correlation $R = 0.2$, $p = 8.1 \times 10^{-9}$ and $R = -0.29$, $p = 6.1 \times 10^{-17}$ respectively), but not associated with any of the IgG glycosylation markers.

I also found IgG-mediated inflammation directly linked with renal function (see below), confirming my previous findings (chapter 5). In this study, IgG glycosylation mediated part of the association of renal function with age as well as its comorbidities, such as decline of

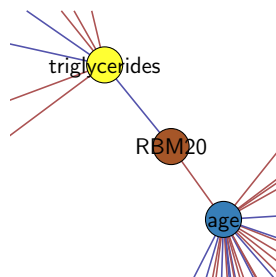
lung function. This supports the theory of ‘inflammageing’, which proposes chronic low-grade inflammation as mechanism that drives disease onset during ageing (Dall’Olio et al., 2013).

9.3.6 IgG-mediated inflammation associated with renal function



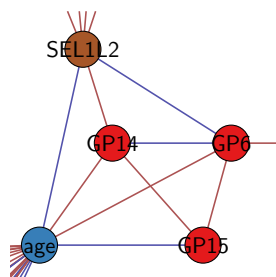
I find the eGFR directly linked to the IgG glycosylation traits GP6 and GP14, which aligns with my previous findings, presented in chapter 5. Here, the IgG glycans partially mediated the association of renal function with age and might thus be a cause of the age-related decline of renal filtration.

9.3.7 RBM20 as mediator of dyslipidaemia with advancing age



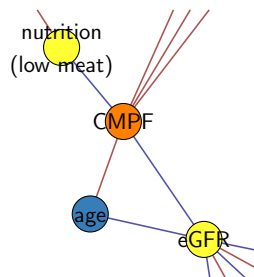
The expression of RBM20 mediates the association between triglyceride levels and age in the Age-MGM. Mutations of the RBM20 gene have been found to cause cardiomyopathy (Li et al., 2010). My model indicates that this association might be due to dyslipidaemia, which is a hallmark of cardiovascular diseases.

9.3.8 SEL1L2 associated with IgG glycosylation



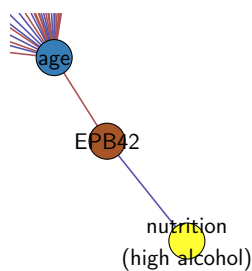
SEL1L2 is directly connected to age and it is the only gene expression marker directly linked to both glycosylation markers GP6 and GP14, which represent the proportion of FA2B and FA2G2 glycans in IgG glycans, respectively. Its function remains elusive, but its paralog, SEL1L, is essential for the degradation of misfolded proteins in the endoplasmic reticulum (Sun et al., 2014) and particularly glycoproteins (Mueller et al., 2008). This possibly underlies the previously reported association between SEL1L and inflammatory bowel disease, particularly Crohn’s disease (Sun et al., 2016), which also is associated with glycosylation of IgG (Theodoratou et al., 2014).

9.3.9 Association of low meat diet and renal function mediated by CMPF



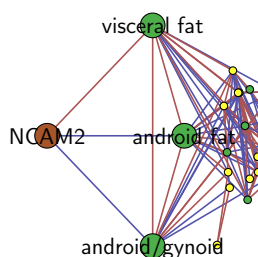
I found the metabolite 3-carboxy-4-methyl-5-propyl-2-furanpropionic acid (CMPF) mediating the association between a low meat diet and renal function. Consistently with this finding, levels of CMPF were reported to increase by fish intake (Hanhineva et al., 2015). It was also shown to be increased in patients suffering from CKD (Niwa, 1996), possibly causing renal function decline (Deguchi et al., 2002). Thus, lower levels of CMPF might explain the beneficial effect of dietary protein restriction on the progress of CKD (Fouque et al., 2001).

9.3.10 Adaption of cell membrane to alcohol intake



EPB42, the only transcriptomics marker in the *LUNG* cluster, is a transmembrane protein that has been associated with spherocytosis and osmotic fragility (Bouhassira et al., 1992). In the Age-MGM it was directly connected to high alcohol intake, mediating its association with age. Consumption of alcohol is known to affect cell membranes and might cause adaption of the cell membranes upon long-term exposure (Hoek et al., 1988). Altered expression of EPB42 might be one of these adaptations.

9.3.11 NCAM2 affects body composition



The gene expression of NCAM2 is the only 'omics' variable in the *FAT* cluster. My model suggests that previous associations of mutations in the NCAM2 gene and waist circumference (Wang et al., 2011a) might be due to a change in its expression level. A recent study found that the knock-out of a homologue gene in mice leads to a reduction of food intake and body weight (Tu et al., 2013), which is possibly related to its effect on olfaction (Cremer et al., 1994). However, the association of NCAM2 expression and body fat distribution appears to be independent of diet in the Age-MGM.

9.4 Conclusion

This is, to my knowledge, the first study combining data from four ‘omics’ technologies and clinical phenotypes using an integrated statistical approach. Despite the relatively small sample size, my model confirmed causal mechanisms of disease, which have been previously found in highly specific experiments and clinical trials, using observational data from a generally healthy cohort only. This illustrates the potential of this approach for future studies. Moreover, I uncovered several new potential mechanisms that might contribute to disease comorbidities. By integrating multiple ‘omics’ datasets, I found the hormone oxytocin as a central mediator that connects inflammation and obesity and, thus, supports the theory of inflammageing. Also, I found urate as key factor connecting body composition and renal function, as well as several phenotypes of the metabolic syndrome.

However, I want to note some limitations of this study. Due to the limited availability of large multi-omics datasets and comprehensive collections of clinical phenotypes, this study is restricted by its relatively small sample size of 510 individuals and, more importantly, I was not able to get access to comparable data from an independent cohort to replicate my findings. For the time being, I could only demonstrate the stability of my results by inferring separate models from two disjoint sets of the same dataset that include only one twin of each twin pair, respectively. Also, all of the participants were female. As a consequence, the model and the conclusions drawn from it might be only partly transferable to the entire population. However, more and larger multi-omics dataset will be available in near future, for instance from the UK Biobank or the US Precision Medicine Initiative, which will facilitate subsequent studies using this multi-omics integration approach.

The limited number of samples also made prior selection of variables indispensable. This selection can be expected to influence the topology and modularity of the final network model. However, in this study, doubling the number of ‘omics’ variables by renouncing prior selection of metabolites did result in a model with very similar topology and module assignments. Again, upcoming larger datasets will help to overcome this limitation by reducing the dimensionality of the data without relying on prior knowledge.

Finally, my approach can detect mediation by variables included in the model and thereby enables differentiation between direct and indirect effects, it does, however, not infer causality. Thus, based on the Age-MGM, I only hypothesize about causal directions. Mendelian randomization might enable inference of causal direction using SNPs as instrumental variables. Much larger sample sizes are needed than available for this study, though. Ideally, potentially causal edges in this model should be further investigated in dedicated functional studies or randomized clinical trials to establish causality and infer causal direction.

Nonetheless, this study highlights the importance and the feasibility of data integration across ‘omics’ layers including phenomics while considering multivariate dependencies. In the future this will help to focus on few, interesting associations, which can then be individually tested in model organisms and clinical trials. Eventually this will speed up drug discovery by excluding irrelevant pathways and potential drug targets early in the development and thus limiting the set of potential targets and reducing costs of drug discovery.

CHAPTER 10

Conclusion and future directions

Here, I summarise my findings and discuss how the individual chapters relate to each other. Then, I discuss some limitations I encountered and give future directions.

In this thesis, I analysed molecular changes associated with biological ageing and particularly kidney disease. Using systems biology methods, I identified several pathways associated with age-related diseases and their comorbidities. I started by carrying out association studies of circulating metabolites and IgG glycans with telomere length and renal function. Next, I assessed the potential use of faecal metabolites to investigate the relationship of the gut microbiome with ageing and renal function. Then, I analysed metabolic multi-fluid processes in CKD patients, before I finally integrated various age-related phenotypes with age-related 'omics' markers to elucidate how molecular changes might mediate age-related comorbidities. In this final chapter, I will first summarise the findings of these analyses, and then discuss their relevance and inter-relationships as well as limitations and problems that I encountered.

10.1 Summary of findings

In **chapter 4**, I looked for association of circulating metabolites with telomere length, a known marker of biological ageing. I identified an involvement of lipid metabolism, particularly PLA2 activity, and oxidative stress, indicated by changes of GGT metabolism, in ageing. GGT is part of the GSH cycle, which is a major pathway of detoxification of ROS.

In chapters 5, 6, and 7, I analysed changes in metabolism and IgG glycosylation patterns associated with kidney disease, an age-related disease. In **chapter 5**, I looked for associations of renal function and changes of the inflammatory potential of IgG, which is modulated by glycosylation. IgG glycans are highly related with chronological age (Kristic et al., 2014) as well as several inflammatory diseases (Vučković et al., 2015; Gornik et al., 2008). This first study of IgG glycosylation and renal function identified alterations of the IgG glycome composition, namely the decrease of galactosylation and sialylation and increase of bisecting GlcNAc, as new inflammatory mechanisms involved in CKD.

Subsequently, in **chapter 6**, I compared metabolomic profiles (mainly lipoprotein composition) associated with renal function in diabetics and non-diabetics. My results suggested several molecular similarities between both conditions, namely impairment of amino acid metabolism, increased proportions of triglycerides, and decreased abundance of HDL cholesterol with decreasing renal function. In contrast, associations of IDL, LDL, and VLDL with renal function were very different between diabetes patients and non-diabetics, possibly due to statin treatment of diabetic individuals. More interestingly, energy metabolism and particularly gluconeogenesis precursors were significantly negatively associated with renal function in non-diabetics, but were not associated with renal function in diabetes cases.

In **chapter 7**, I used metabolomics measurements from faeces as a functional readout of the gut microbiome to explore its associations with renal function and age. While the faecal metabolome was heritable and highly associated with the gut microbial community, it was only moderately associated with age and renal function. As previous studies reported changes of the microbiome in late age and with frailty, the lack of significant associations in my analysis might be due to the study population being too young and healthy or due to a lack of power. In contrast, the faecal metabolome was strongly correlated with obesity and visceral fat mass. These latter studies were particularly novel.

In **chapter 8**, I aimed to further investigate metabolic processes of renal disease, which I and others had observed before, by modelling metabolic networks for renal disease patients and controls. Integrating three different fluids – plasma, urine, and saliva – facilitated the investigation of transport and excretion processes. This analysis highlighted steroid metabolism and xanthine catabolism as central processes in kidney disease. Also, differences in the metabolism of salivary amino acids between renal disease patients and healthy controls suggested a shift of the oral microbiome in the disease, possibly related to the comorbidity of renal and periodontal disease.

Finally, in **chapter 9**, I investigated how age-related changes on the molecular level relate to each other and how these changes mediate comorbidities of age-related diseases. To this end, I integrated a selection of age-related ‘omics’ markers with a broad collection of clinical phenotypes. My results highlight several potential mechanisms that might cause such comorbidities, for instance renal function mediating associations of most circulating metabolites with chronological age.

While I focused on a particular aspect of biological ageing using different technologies and analysis methods in each chapter, these studies and their results are highly interrelated. In the following, I first highlight kidney-specific results that I obtained across my studies (section 10.1.1). In a second section (section 10.1.2) I review my results more generally with respect to biological ageing.

10.1.1 Kidney disease

Kidney disease is a common disease, which is increasingly prevalent due to population ageing and the global obesity epidemic, which are the main risk factors of CKD. As general markers are very insensitive, better understanding the pathomechanisms of diabetic and non-diabetic nephropathies is essential to personalise healthcare.

I found lipid profiles strongly associated with renal function in non-diabetic individuals, but not in type 2 diabetes patients. While the commonly used total LDL levels did not show a significant correlation with the eGFR, several of its subclasses did. Similarly, HDL subclasses showed much stronger associations with renal function than total HDL, particularly the proportion of phospholipids in HDL, which also predicted longitudinal change of eGFR. In contrast, changes of amino acid levels were correlated with renal function consistently between diabetic and non-diabetic individuals. Despite the very different platforms used in chapters 6 and 8, concordant changes were observed in both cohorts. Additionally, differential network analysis highlighted the catabolism of xanthines and steroid metabolism. Interestingly, the metabolites that were most strongly correlated with renal function cross-sectionally were neither the most predictive for longitudinal change of renal function in chapter 6, nor the most affected metabolic processes in chapter 8. Combining metabolomics measures with other ageing markers in chapter 9 suggested that the decline of renal function mediates most of the metabolic changes that occur with advancing age.

Looking at IgG glycosylation opened a new perspective on kidney disease compared to its better-established relationship with metabolism. While the immune system and inflammatory processes are known to be involved in the development of CKD (Imig et al., 2013), no study had investigated IgG glycosylation, which modulates its inflammatory potential, before. The association study highlighted several glycosylation features – decreased galactosylation and sialylation, and increased bisecting GlcNAc – that were significantly associated with decreased eGFR. Changes of the immune system and the metabolism are known to be related, for instance due to the increase energy demand in active lymphocytes (Fox et al., 2005). Moreover, the increased availability of mannose was shown to increase the mannose content in IgG glycans (Slade et al., 2016). Though, in my network analysis, most associations of metabolite levels and IgG glycans were mediated by renal function, suggesting that renal disease is associated with both metabolites and glycans independently. Interestingly, this did not apply to a group of steroid hormones, including DHEA-S, which is one of the strongest predictors of age.

DHEA-S was directly associated with IgG but only indirectly with renal function, suggesting an association of steroid metabolism and IgG glycosylation independently of renal function. Steroid metabolism also appeared to be one of the key factors of metabolic disruption in kidney disease in chapter 8. Future studies on the relationship of IgG glycosylation, steroid metabolism, and renal disease might enable the full elucidation of the causal directions of these associations. To further assess their potential as clinical markers of renal disease and more generally unhealthy ageing, longitudinal analyses will be needed. Also, analysis of IgG glycosylation patterns in a clinical cohort with advanced renal disease will be necessary to confirm the associations that I observed here in the general population and to investigate their associations with complications of renal failure.

10.1.2 Biological ageing

I identified metabolic markers of oxidative stress and lipid metabolism associated with biological ageing, measured as telomere length. These are in line with established theories on the cause of biological ageing. Integrating age-related clinical and molecular phenotypes highlighted mechanisms on how such molecular changes potentially cause disease comorbidities. For instance, most age-related changes of the metabolism were mediated by the decline of renal function. These metabolic changes, such as hyperuricemia, mediated the comorbidity of renal disease and changes of body composition, particularly abdominal obesity. In contrast, changes of steroid levels were indirectly associated with renal function, mediated by age.

My network approach also demonstrated strong interdependencies between common age-related diseases independently of age itself. For instance, the expression of the hormone oxytocin mediated the association of body composition with inflammatory IgG. Taking this complex comorbidity structure into account is essential to identify biomarkers that are specific to one disease. For instance, IgG glycosylation was directly connected to several disease markers, including renal function, lung function, and arthritis and, while considered a general, useful marker of ageing, is likely not specific enough to be used as clinical biomarker for either of these individual diseases. Moreover, with the growing interest in personalised medicine it becomes increasingly important to consider the molecular changes and comorbidities of diseases more holistically rather than relying on a single biomarker regardless of a patient's medical history.

10.2 Limitations

While my results highlighted several interesting aspects of biological ageing and age-related comorbidities, there are several problems that I encountered during my work on this thesis. In the following, I will discuss such general problems and limitations.

10.2.1 Systems biology

While systems biology is meant to improve more conventional association studies by offering greater insights in the mechanisms of a phenotype, it comes with its own inherent problems.

The biggest of these problems – in my opinion – is the necessity for complete data matrices. Particularly in a cohort like TwinsUK, which evolved over time, each piece of data is available for different subsets of the participants and measured at different points in time. Combining multiple of such datasets inevitably leads to smaller and smaller datasets, thus not providing sufficient power. This was the reason for choosing a correlation network rather than a more integrative approach to combine the associations of telomere length in chapter 4, and it was also the reason for the relatively small sample size underlying the combined ageing network in chapter 9. Larger studies that are designed to collect multi-omics data, such as the UK Biobank, will eventually solve this problem. However, even then it will probably not be trivial to obtain all the phenotypes of interest.

Another problem of systems biology, and particularly large networks, is their sometimes overwhelming complexity. While association studies are usually rather straight-forward to conduct – fitting regression models, replicating, and then discussing the strongest associations – there are a multitude of choices to be made in each step of a systems biology project, each having a major impact on the final result and making standardisation difficult.

Starting from network inference, when using penalised approaches, such as the very common graphical LASSO, the regularisation parameter has to be selected. There are several strategies for choosing the optimal regularisation parameter, for instance by optimisation of the AIC or the BIC, cross-validation, or stability-based approaches. Graphs inferred with the same algorithm – the graphical LASSO – but different regularisation parameters can vary by orders of magnitude, thus changing the structure and interpretation of networks. The GeneNet algorithm is a great alternative, avoiding costly permutation procedures and controlling false positive discoveries.

Another problem of network biology is the visualisation of results. While the network layout does not have a direct effect on network topology measures, it is a major determinant of how we perceive a network, thus changing the intuitive understanding. Again, there are a variety of algorithms to compute the ‘optimal’ layout and each of them produces dramatically different pictures for the same graph. Increasing availability of three-dimensional visualisation techniques, such as 3D screens, 3D printing, and eventually holograms, will facilitate more accessible visualisation of large networks.

Finally, it is very hard to replicate networks across different cohorts. The first problem is – again – the limited availability of data, which is particularly the case for multi-omics and phenomics networks. However, the more complex issue is the comparison of two networks. For an unweighted network one can just count the overlap of edges between two networks. However, all networks have some continuous statistic underlying the binary absence or presence of an edge. Comparing these underlying statistics, such as shrinkage partial correlations, is not straight-forward as they depend on the amount of regularisation, which in turn depends on sample size and the correlation structure within the data. One could use a differential network approach as described in chapter 8 to identify significantly different edges and conclude concordance between cohorts for all other edges. However, this would require sharing large amounts of raw data between cohorts, which is often not feasible. In the meantime, splitting the available data and use subsets for validation, as I did in chapter 9, is a pragmatic way to demonstrate the stability of the findings, at least within one cohort.

Apart from these technical limitations all systems biology approaches are oversimplifications of the real system, often ignoring cell compartmentalisation, cell types, inter-cell communication, tissues, organs, etc. Inferring an actual holistic model of a cell, not to speak of complex disease processes, are far-fetched goals. Thus, all systems biology projects need to be reductionist to some extent.

Despite these limitations, I demonstrated the feasibility of integrative network approaches in this thesis. With time and increasing application of systems biology methods, best practices will evolve and further simplify the choice of the ‘right’ algorithms, thus facilitating integration of larger datasets. Also, growing datasets as well as increasing computational power will help

to overcome several of the limitations, such as the lack of replication datasets, and facilitate permutation based approaches to quantify the stability of large networks. While this will facilitate future studies, I think that some of the complexity of systems biology arises from its aim to fully understand biological processes, which are inherently complex. This increased complexity and the problems it brings should be acknowledged and be considered the price for the stronger conclusions.

10.2.2 Causality

The main aim of systems biology is understanding disease mechanisms, which implies causality of the associations. The lack of causal inference is one of the limitations of this work. However, this limitation is shared by most epidemiological research that exclusively relies on observational data. Thus, based on my findings, I could only draw hypothesis on underlying mechanisms. However, I believe that systems biology does help to create stronger hypotheses, by utilising data from multiple sources and considering more potential confounding factors. The identification of direct, independent associations and more complex correlation structures allows us to draw hypotheses on the underlying mechanisms, even though causality itself has to be further investigated in subsequent experiments.

Furthermore, in some cases there might not even be a single causal direction. For instance, metabolic reactions are often equilibrium reactions, which work both ways. Moreover, even if a metabolic reaction is irreversible, accumulation of the product might change the reaction kinetics, which will lead to slower reaction rates and potentially also accumulation of the educt. Consequently, even though there is a true causal direction from educt to product, the increased concentration of the product can be considered the cause of the increased concentration of the educt under some circumstances. Thus, causal direction depends on the given conditions, and might change for instance in disease states.

10.2.3 Translation to clinical application

While causality cannot be proven in many cases, it is not strictly required for some applications. For instance, a biomarker does not necessarily have to be causal for a disease. In contrast, a good biomarker should be directly connected to the disease to allow for the specificity needed in clinical settings. For instance, accumulation of creatinine is not a cause but a consequence of kidney disease, it is, however, directly related to renal function. Similarly, the concentration of albumin in urine itself is not harmful, but a mere marker of renal damage.

Nevertheless, biomarker discovery is challenging even without the need for causality. While IgG glycosylation was strongly associated with eGFR, it did not classify CKD patients more accurately than a rather primitive model based on age, sex, and BMI in chapter 5. This limitation is not specific to this thesis. For example, Sekula et al. (2016) found several metabolites very strongly correlated with eGFR ($p < 1.0 \times 10^{-100}$), however, did not improve the classification of

CKD cases over an equally primitive model. Defining kidney disease based on creatinine measures is probably the biggest limitation for the discovery of new biomarkers, as – by definition – no biomarker can classify cases and controls more accurately than creatinine. Also, training predictive models based on this inaccurate gold-standard might lead to underperformance of new, possibly superior biomarkers.

More important than classification performance is often the ability to predict longitudinal change, which facilitates decisions on intervention strategies. However, the compounds that are most strongly associated with a disease cross-sectionally are not necessarily the best longitudinal predictors. For instance, creatinine is used to classify renal disease patients but is not able to differentiate between fast and slow disease progression. In contrast, the proportion of phospholipids in HDL predicted longitudinal change of eGFR in chapter 6, even though it was only moderately correlated cross-sectionally.

While identification of reliable biomarkers is an important issue, for instance to improve management of renal disease, it is not the primary aim of network-based approaches. While these multivariate approaches and the resulting hypotheses on mechanisms of disease are less straight-forward to implement in clinics, they are equally important. For instance, in chapter 8, I identified xanthine metabolism among the processes most affected by renal disease. As many drugs are metabolised by this pathway, renal function should be carefully considered prior to drug treatment, even for independent diseases. Also, in chapter 9 I identified urate as potential mediator between renal disease and obesity and oxytocin as potential mediator of age-related obesity. Knowing these mediating factors facilitates targeted treatment and might help to avoid such comorbidities in the future. However, additional studies on clinical samples must further clarify these pathways before they can be exploited in clinical care.

10.3 Other confounding factors

Even if a biomarker candidate performs well in disease classification and prediction, it cannot be directly translated into clinical application. Most of the analyses in this thesis were based on the generally healthy population and findings are not always transferable to more diseased individuals. For instance, the amount of phospholipids in HDL only predicted change in eGFR in non-diabetics but not in diabetics. Similarly, other biomarkers or disease processes might differ between age groups, genders, ethnicities, or even between twin pairs and singletons. Replication of results across independent cohorts facilitates the identification for instance of ethnicity-specific markers, and helped here to identify differences between diabetic and non-diabetic individuals. However, further replication of my results in cohorts with a more diverse genetic background is needed to confirm the transferability of my results. Also, disease processes might differ considerably in children or the elderly as well as patients in clinical settings, who might receive various treatments and have multiple interrelated diseases. Consequently, generalisation of my finding to these population groups needs to be examined in future studies.

10.4 Future directions and lessons learnt

My network analysis indicated independent associations of metabolites and IgG glycans with renal function. Hence, combining both in future analyses might improve the classification of CKD patients. Similarly, the MS and NMR platforms that were used for metabolomics measurements mostly detect different metabolites. Future studies should investigate if they capture complementary information on renal disease. Using a graphical model approach to combine all three datasets with renal function could identify independent associations and might unveil new disease processes.

Moreover, upcoming larger datasets will address several of the aforementioned limitations. Using a knowledge-driven prior selection of ‘omics’ variables allowed me to infer a multi-omics model in chapter 9, however, I possibly missed mechanisms of less age-related compounds. Larger datasets will facilitate a more systematic selection of variables or dispensing with prior selection altogether. As several very large projects – the American precision medicine initiative, the British UK Biobank, and the German national cohort – extensively profile large cohorts of up to 500,000 people, replication of such integrative studies will become more feasible. Moreover, these large studies with genotyping data available will facilitate Mendelian randomisation studies, which could help to understand causal mechanisms underlying the cross-sectional associations.

However, the increasing amount of data comes with its own challenges. For instance, visualisation is problematic for graphs with some hundred nodes, but will be extremely difficult for several thousand nodes. Also, the choice of regularisation parameters in approaches such as the graphical LASSO often relies on computationally expensive permutation-based procedures. Particularly differential models, which would require nested permutations, are computationally demanding. Developing novel methods and computationally efficient solutions to these problems as well as allowing for parallel implementation will be crucial. Thus, pioneering studies, integrating high-dimensional data from multiple sources or analysing network differences, not only help to identify important disease mechanisms, but also reveal the limitations of current systems biology methods and set the ground for future studies.

Epidemiological research can, regardless of sample size, only provide candidate biomarkers and generate hypotheses on mechanisms of disease. Wet-lab experiments will have to confirm mechanisms proposed by dry-lab studies to enable translation of the results into clinical practice. With ever-growing data dimensionality, increasing the number of both variables and samples, association studies will find more significant correlations than wet-labs will ever be able to further investigate. By uncovering the compounds most closely related to a trait, systems biology has the potential to triage the compounds that are most promising for further experiments.

In conclusion, I identified putative mechanisms of ageing and age-related disease comorbidities using systems biology methods. These highlight the potential of this field to uncover disease processes and assist the new science of personalised or precision medicine.

Bibliography

- Adamczyk, B., Tharmalingam, T. and Rudd, P. M. (2012). 'Glycans as cancer biomarkers.' *Biochimica et biophysica acta* 1820.9, pp. 1347–53.
- Agarwala, A. P., Rodrigues, A., Risman, M., McCoy, M., Trindade, K. et al. (2015). 'High-Density Lipoprotein (HDL) Phospholipid Content and Cholesterol Efflux Capacity Are Reduced in Patients With Very High HDL Cholesterol and Coronary Disease'. *Arteriosclerosis, Thrombosis, and Vascular Biology* 35.6, pp. 1515–1519.
- Albert, R., Jeong, H. and Barabási, A.-L. (2000). 'Error and attack tolerance of complex networks'. *Nature* 406.6794, pp. 378–382.
- Alegria-Torres, J. A., Baccarelli, A. and Bollati, V. (2011). 'Epigenetics and lifestyle'. *Epigenomics* 3.3, pp. 267–277.
- Allocco, D. J., Kohane, I. S. and Butte, A. J. (2004). 'Quantifying the relationship between co-expression, co-regulation and gene function.' *BMC bioinformatics* 5.1, p. 18.
- Allsopp, R. C., Chang, E., Kashefi-Azham, M., Rogaev, E. I., Piatyszek, M. A. et al. (1995a). 'Telomere Shortening Is Associated with Cell Division in Vitro and in Vivo'. *Experimental Cell Research* 220.1, pp. 194–200.
- Allsopp, R. C. and Harley, C. B. (1995b). 'Evidence for a Critical Telomere Length in Senescent Human Fibroblasts'. *Experimental Cell Research* 219.1, pp. 130–136.
- Altieri, P., Sau, G., Cao, R., Barracca, A., Menneas, A. et al. (2002). 'Immunosuppressive treatment in dialysis patients.' *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 17 Suppl 8, pp. 2–9.
- Antharam, V. C., McEwen, D. C., Garrett, T. J., Dossey, A. T., Li, E. C. et al. (2016). 'An integrated metabolomic and microbiome analysis identified specific gut microbiota associated with fecal cholesterol and coprostanol in *Clostridium difficile* infection'. *PLoS ONE* 11.2, pp. 1–23.
- Anthony, R. M., Nimmerjahn, F., Ashline, D. J., Reinhold, V. N., Paulson, J. C. et al. (2008). 'Recapitulation of IVIG Anti-Inflammatory Activity with a Recombinant IgG Fc'. *Science* 320.5874, pp. 373–376.
- Anthony, R. M. and Ravetch, J. V. (2010). 'A Novel Role for the IgG Fc Glycan: The Anti-inflammatory Activity of Sialylated IgG Fcs'. *Journal of Clinical Immunology* 30.S1, pp. 9–14.
- Aozaki, S. (1989). 'Decreased membrane fluidity in erythrocytes from patients with Crohn's disease'. *Gastroenterol.Jpn.* 24.3, pp. 246–254.
- Apro, J., Tietge, U. J. F., Dijkers, A., Parini, P., Angelin, B. et al. (2016). 'Impaired Cholesterol Efflux Capacity of High-Density Lipoprotein Isolated From Interstitial Fluid in Type 2 Diabetes Mellitus-Brief Report.' *Arteriosclerosis, thrombosis, and vascular biology* 36.5, pp. 787–91.
- Araújo, M. V. F., Hong, B.-Y., Fava, P. L., Khan, S., Burleson, J. A. et al. (2015). 'End stage renal disease as a modifier of the periodontal microbiome.' *BMC nephrology* 16, p. 80.
- Arca, M., Montali, A., Valiante, S., Campagna, F., Pigna, G. et al. (2007). 'Usefulness of Atherogenic Dyslipidemia for Predicting Cardiovascular Risk in Patients With Angiographically Defined Coronary Artery Disease'. *American Journal of Cardiology* 100.10, pp. 1511–1516.
- Ariyamuthu, V. K., Nolph, K. D. and Ringdahl, B. E. (2013). 'Periodontal disease in chronic kidney disease and end-stage renal disease patients: a review.' *Cardiorenal medicine* 3.1, pp. 71–8.
- Arlt, W. and Stewart, P. M. (2005). 'Adrenal corticosteroid biosynthesis, metabolism, and action'. *Endocrinology and Metabolism Clinics of North America* 34.2, pp. 293–313.
- Arnold, J., Dai, J., Nahapetyan, L., Arte, A., Johnson, M. A. et al. (2010). 'Predicting successful aging in a population-based sample of georgia centenarians.' *Current gerontology and geriatrics research* 2010, p. 9.
- Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. and Kastenmüller, G. (2015). 'SNiPA: an interactive, genetic variant-centered annotation browser.' *Bioinformatics (Oxford, England)* 31.8, pp. 1334–6.
- Artandi, S. E. and DePinho, R. a. (2009). 'Telomeres and telomerase in cancer'. *Carcinogenesis* 31.1, pp. 9–18.
- Arthur, J. M. and Klein, J. B. (2010). 'Proteomics in CKD.' *Advances in chronic kidney disease* 17.6, pp. 453–4.

- Ayodele, O. E. and Alebiosu, C. O. (2010). 'Burden of Chronic Kidney Disease: An International Perspective'. *Advances in Chronic Kidney Disease* 17.3, pp. 215–224.
- Bacalini, M. G., Friso, S., Olivieri, F., Pirazzini, C., Giuliani, C. et al. (2014). 'Present and future of anti-ageing epigenetic diets'. *Mechanisms of ageing and development* 136–137, pp. 101–15.
- Balakrishnan, V. S., Coles, G. A. and Williams, J. D. (1993). 'A potential role for endogenous adenosine in control of human glomerular and tubular function'. *The American journal of physiology* 265.4 Pt 2, F504–10.
- Barabási, A.-L. and Oltvai, Z. N. (2004). 'Network biology: understanding the cell's functional organization'. *Nature reviews. Genetics* 5.2, pp. 101–13.
- Barrios, C., Beaumont, M., Pallister, T., Villar, J., Goodrich, J. K. et al. (2015). 'Gut-Microbiota-Metabolite Axis in Early Renal Function Decline'. *PLOS ONE* 10.8. Ed. by Remuzzi, G., e0134311.
- Barrios, C., Spector, T. D. and Menni, C. (2016). 'Blood, urine and faecal metabolite profiles in the study of adult renal disease'. *Archives of Biochemistry and Biophysics* 589, pp. 81–92.
- Barrios, C., Zierer, J. et al. (2016). 'Glycosylation Profile of IgG in Moderate Kidney Dysfunction'. *Journal of the American Society of Nephrology* 27.3, pp. 933–941.
- Bartz, J. A., Zaki, J., Bolger, N. and Ochsner, K. N. (2011). 'Social effects of oxytocin in humans: context and person matter'. *Trends in cognitive sciences* 15.7, pp. 301–9.
- Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software* 67.1, p. 51.
- Baulieu, E. E., Thomas, G., Legrain, S., Lahlou, N., Roger, M. et al. (2000). 'Dehydroepiandrosterone (DHEA), DHEA sulfate, and aging: contribution of the DHEAge Study to a sociobiomedical issue.' *Proceedings of the National Academy of Sciences of the United States of America* 97.8, pp. 4279–4284.
- Baumann, U. and Schmidt, R. E. (2001). 'The role of Fc receptors and complement in autoimmunity.' *Advances in experimental medicine and biology* 495.6, pp. 219–25.
- Beaumont, M., Goodrich, J. K., Jackson, M. A., Yet, I., Davenport, E. R. et al. (2016). 'Heritable components of the human fecal microbiome are associated with visceral fat'. *Genome Biology* 17.1, p. 189.
- Beck, L. H., Bonegio, R. G. B., Lambeau, G., Beck, D. M., Powell, D. W. et al. (2009). 'M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy'. *The New England journal of medicine* 361.1, pp. 11–21.
- Becker, N., Kunath, J., Loh, G. and Blaut, M. (2011). 'Human intestinal microbiota: Characterization of a simplified and stable gnotobiotic rat model'. *Gut Microbes* 2.1, pp. 25–33.
- Beckman, K. B. and Ames, B. N. (1998). 'The free radical theory of aging matures'. *Physiological reviews* 78.2, pp. 547–81.
- Beisser, D., Klau, G. W., Dandekar, T., Müller, T. and Dittrich, M. T. (2010). 'BioNet: an R-Package for the functional analysis of biological networks'. *Bioinformatics (Oxford, England)* 26.8, pp. 1129–30.
- Bell, J. T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J. et al. (2012). 'Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population.' *PLoS genetics* 8.4, e1002629.
- Bell, R., Hubbard, A., Chettier, R., Chen, D., Miller, J. P. et al. (2009). 'A human protein interaction network shows conservation of aging processes between human and invertebrate species.' *PLoS genetics* 5.3, e1000414.
- Berchtold, N. C., Cribbs, D. H., Coleman, P. D., Rogers, J., Head, E. et al. (2008). 'Gene expression changes in the course of normal brain aging are sexually dimorphic.' *Proceedings of the National Academy of Sciences of the United States of America* 105.40, pp. 15605–10.
- Berg, J. M., Tymoczko, J. and Stryer, L. (2002). *Biochemistry*. New York: W H Freeman, p. 1050.
- Biagi, E., Candela, M., Fairweather-Tait, S., Franceschi, C. and Brigidi, P. (2012). 'Aging of the human metaorganism: the microbial counterpart.' *Age (Dordrecht, Netherlands)* 34.1, pp. 247–67.
- Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L. et al. (2010). 'Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians.' *PloS one* 5.5, e10667.
- Bingham, S. A., Welch, A. A., McTaggart, A., Mulligan, A. A., Runswick, S. A. et al. (2001). 'Nutritional methods in the European Prospective Investigation of Cancer in Norfolk'. *Public Health Nutrition* 4.03, p. 847.
- Blasco, M. (2007). 'Telomere length, stem cells and aging'. *Nature chemical biology* 3.10, pp. 640–9.
- Bochud, M., Elston, R. C., Maillard, M., Bovet, P., Schild, L. et al. (2005). 'Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean)'. *Kidney international* 67.1, pp. 61–9.
- Boero, R., Pignataro, A., Rollino, C. and Quarello, F. (2000). 'Do corticosteroids improve survival in acute renal failure due to cholesterol atheroembolism?' *Nephrology Dialysis Transplantation* 15.3, p. 441.
- Boirie, Y., Albright, R., Bigelow, M. and Nair, K. S. (2004). 'Impairment of phenylalanine conversion to tyrosine in end-stage renal disease causing tyrosine deficiency'. *Kidney International* 66.2, pp. 591–596.

- Bolasco, P., Caria, S., Cupisti, A., Secci, R. and Saverio Dioguardi, F. (2011). 'A novel amino acids oral supplementation in hemodialysis patients: a pilot study.' *Renal failure* 33.1, pp. 1–5.
- Bonnet, S., Dumas-de-La-Roque, E., Bégueret, H., Marthan, R., Fayon, M. et al. (2003). 'Dehydroepiandrosterone (DHEA) prevents and reverses chronic hypoxic pulmonary hypertension.' *Proceedings of the National Academy of Sciences of the United States of America* 100, pp. 9488–9493.
- Borenstein, M., Hedges, L. V., Higgins, J. P. and Rothstein, H. R. (2010). 'A basic introduction to fixed-effect and random-effects models for meta-analysis'. *Research Synthesis Methods* 1.2, pp. 97–111.
- Boron, W. F. and Boulpaep, E. L. (2012). *Medical Physiology: a cellular and molecular approach*. Ed. by O'Grady, E., Hall, A., Tannian, P., Casey, J. and Stave, S. 2nd. Philadelphia, PA: Saunders Elsevier, p. 1352.
- Bouhassira, E. E., Schwartz, R. S., Yawata, Y., Ata, K., Kanzaki, A. et al. (1992). 'An alanine-to-threonine substitution in protein 4.2 cDNA is associated with a Japanese form of hereditary hemolytic anemia (protein 4.2NIPPON)'. *Blood* 79.7, pp. 1846–1854.
- Bratic, A. and Larsson, N.-g. (2013). 'The role of mitochondria in aging'. *The Journal of clinical investigation* 123.3, pp. 951–7.
- Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. and Brenner, H. (2011). 'Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication'. *The American Journal of Human Genetics* 88.4, pp. 450–457.
- Brink, D. M. van den and Wanders, R. J. A. (2006). 'Phytanic acid: production from phytol, its breakdown and role in human disease'. *Cellular and molecular life sciences : CMLS* 63.15, pp. 1752–65.
- Brion, M.-j. A., Benyamin, B., Visscher, P. M. and Smith, G. D. (2014). 'Beyond the Single SNP : Emerging Developments in Mendelian Randomization in the " Omics " Era'. *Current Epidemiology Reports* 1.4, pp. 228–236.
- Brouillette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. and Samani, N. J. (2003). 'White cell telomere length and risk of premature myocardial infarction'. *Arteriosclerosis, Thrombosis, and Vascular Biology* 23.5, pp. 842–846.
- Buuren, S. van and Groothuis-Oudshoorn, K. (2011). 'mice : Multivariate Imputation by Chained Equations in R'. *Journal of Statistical Software* 45.3, pp. 1–67.
- Calado, R. T. and Young, N. S. (2009). 'Telomere Diseases'. *New England Journal of Medicine* 361.24, pp. 2353–2365.
- Callahan, M. M., Robertson, R. S., Arnaud, M. J., Branfman, A. R., McComish, M. F. et al. (1982). 'Human metabolism of [1-methyl-14C]- and [2-14C]caffeine after oral administration'. *Drug metabolism and disposition: the biological fate of chemicals* 10.4, pp. 417–23.
- Calverley, P. M. A., Rabe, K. F., Goehring, U.-M., Kristiansen, S., Fabbri, L. M. et al. (2009). 'Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials'. *Lancet (London, England)* 374.9691, pp. 685–94.
- Cangelosi, G. A. and Meschke, J. S. (2014). 'Dead or alive: molecular assessment of microbial viability'. *Applied and environmental microbiology* 80.19, pp. 5884–91.
- Cano, N. J. M., Fouque, D. and Leverve, X. M. (2006). 'Application of branched-chain amino acids in human pathological states: renal failure'. *The Journal of nutrition* 136.1 Suppl, 299S–307S.
- Cantin, B., Brun, L. D., Gagné, C., Murthy, M. R., Lupien, P. J. et al. (1992). 'Alterations in erythrocyte membrane lipid composition and fluidity in primary lipoprotein lipase deficiency'. *Biochimica et biophysica acta* 1139.1–2, pp. 25–31.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D. et al. (2010). 'QIIME allows analysis of high-throughput community sequencing data'. *Nature Methods* 7.5, pp. 335–336.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Lozupone, C. A. et al. (2011). 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample'. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl. Supplement_1, pp. 4516–22.
- Caprari, P., Scuteri, A., Salvati, A. M., Bauco, C., Cantafora, A. et al. (1999). 'Aging and red blood cell membrane: A study of centenarians'. *Experimental Gerontology* 34.1, pp. 47–57.
- Carr, D. B., Utzschneider, K. M., Hull, R. L., Kodama, K., Retzlaff, B. M. et al. (2004). 'Intra-Abdominal Fat Is a Major Determinant of the National Cholesterol Education Program Adult Treatment Panel III Criteria for the Metabolic Syndrome'. *Diabetes* 53.8, pp. 2087–2094.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H. et al. (2014). 'The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases'. *Nucleic acids research* 42.Database issue, pp. D459–71.
- Cassman, M. (2005). 'Barriers to progress in systems biology'. *Nature* 438.7071, p. 1079.
- Cawthon, R. M. (2002). 'Telomere measurement by quantitative PCR'. *Nucleic acids research* 30.10, e47.
- Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. and Kerber, R. A. (2003). 'Association between telomere length in blood and mortality in people aged 60 years or older'. *Lancet* 361.9355, pp. 393–395.
- Cevenini, E., Invidia, L., Lescai, F., Salvio, S., Tieri, P. et al. (2008). 'Human models of aging and longevity'. *Expert opinion on biological therapy* 8.9, pp. 1393–405.

- Chaves, V. E., Tilelli, C. Q., Brito, N. A. and Brito, M. N. (2013). 'Role of oxytocin in energy metabolism'. *Peptides* 45, pp. 9–14.
- Chen, S., Witten, D. M. and Shojaie, A. (2015). 'Selection and estimation for mixed graphical models'. *Biometrika* 102.1, pp. 47–64.
- Choi, K. C., Lee, J., Moon, K. H., Park, K. K., Kim, S. W. et al. (1993). 'Chronic caffeine ingestion exacerbates 2-kidney, 1-clip hypertension and ameliorates deoxycorticosterone acetate-salt hypertension in rats.' *Nephron* 65.4, pp. 619–22.
- Choo, J. M., Leong, L. E. X. and Rogers, G. B. (2015). 'Sample storage conditions significantly influence faecal microbiome profiles.' *Scientific reports* 5, p. 16350.
- Civelek, M. and Lusis, A. J. (2014). 'Systems genetics approaches to understand complex traits.' *Nature reviews. Genetics* 15.1, pp. 34–48.
- Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., Weerd, H. de et al. (2011). 'Composition, variability, and temporal stability of the intestinal microbiota of the elderly.' *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl, pp. 4586–91.
- Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M. et al. (2012). 'Gut microbiota composition correlates with diet and health in the elderly.' *Nature* 488.7410, pp. 178–84.
- Clarke, G., Stilling, R. M., Kennedy, P. J., Stanton, C., Cryan, J. F. et al. (2014). 'Gut Microbiota: The Neglected Endocrine Organ.' *Molecular endocrinology* 28.8, pp. 1221–1238.
- Clauset, A., Newman, M. E. J. and Moore, C. (2004). 'Finding community structure in very large networks'. *Physics* 70.6, pp. 1–6.
- Codd, V., Mangino, M., Harst, P. van der, Braund, P. S., Kaiser, M. et al. (2010). 'Common variants near TERC are associated with mean telomere length.' *Nature genetics* 42.3, pp. 197–199.
- Codd, V., Nelson, C. P., Albrecht, E., Mangino, M., Deelen, J. et al. (2013). 'Identification of seven loci affecting mean telomere length and their association with disease'. *Nature Genetics* 45.4, pp. 422–427.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd. Hillsdale, New Jersey: Routledge, p. 567.
- Collino, S., Montoliu, I., Martin, F. P. J., Scherer, M., Mari, D. et al. (2013). 'Metabolic Signatures of Extreme Longevity in Northern Italian Centenarians Reveal a Complex Remodeling of Lipids, Amino Acids, and Gut Microbiota Metabolism'. *PLoS ONE* 8.3, pp. 1–12.
- Collins, K. and Mitchell, J. R. (2002). 'Telomerase in the human organism.' *Oncogene* 21.4, pp. 564–579.
- Cook, H. T. (2013). 'Complement and kidney disease'. *Current Opinion in Nephrology and Hypertension* 22.3, pp. 295–301.
- Cook, V. J., Turenne, C. Y., Wolfe, J., Pauls, R. and Kabani, A. (2003). 'Conventional methods versus 16S ribosomal DNA sequencing for identification of nontuberculous mycobacteria: cost analysis.' *Journal of clinical microbiology* 41.3, pp. 1010–5.
- Coresh, J., Selvin, E., Stevens, L. a., Manzi, J., Kusek, J. W. et al. (2007). 'Prevalence of chronic kidney disease in the United States.' *JAMA* 298.17, pp. 2038–47.
- Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I. et al. (2009). 'Bacterial community variation in human body habitats across space and time.' *Science (New York, N.Y.)* 326.5960, pp. 1694–7.
- Coulson, R. and Scheinman, S. J. (1989). 'Xanthine effects on renal proximal tubular function and cyclic AMP metabolism.' *The Journal of pharmacology and experimental therapeutics* 248.2, pp. 589–95.
- Coutinho, A. E. and Chapman, K. E. (2011). 'The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights'. *Molecular and Cellular Endocrinology* 335.1, pp. 2–13.
- Cremer, H., Lange, R., Christoph, A., Plomann, M., Vopper, G. et al. (1994). 'Inactivation of the N-CAM gene in mice results in size reduction of the olfactory bulb and deficits in spatial learning.' *Nature* 367.6462, pp. 455–459.
- Cryan, J. F. and Dinan, T. G. (2012). 'Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour.' *Nature reviews. Neuroscience* 13.10, pp. 701–12.
- Csardi, G. and Nepusz, T. (2006). 'The igraph software package for complex network research'. *InterJournal Complex Sy*, p. 1695.
- Dall'Olio, F., Vanhooren, V., Chen, C. C., Slagboom, P. E., Wuhrer, M. et al. (2013). 'N-glycomic biomarkers of biological aging and longevity: A link with inflammaging'. *Ageing Research Reviews* 12.2, pp. 685–698.
- Darshi, M., Van Espen, B. and Sharma, K. (2016). 'Metabolomics in Diabetic Kidney Disease: Unraveling the Biochemistry of a Silent Killer'. *American Journal of Nephrology* 44.2, pp. 92–103.
- D'Aspremont, A., Banerjee, O. and Ghaoui, L. E. (2006). 'First-order methods for sparse covariance selection'. *Journal of machine learning research : JMLR* 10, pp. 883–906.
- Davies, G., Harris, S. E., Reynolds, C. a., Payton, A., Knight, H. M. et al. (2014). 'A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing.' *Molecular psychiatry* 19.1, pp. 76–87.

- De Grey, A. D. (2005). 'Reactive Oxygen Species Production in the Mitochondrial Matrix: Implications for the Mechanism of Mitochondrial Mutation Accumulation'. *Rejuvenation Research* 8.1, pp. 13–17.
- De Lange, T. (2005). 'Shelterin: The protein complex that shapes and safeguards human telomeres'. *Genes and Development* 19.18, pp. 2100–2110.
- De Palencia, P. F., Fernández, M., Mohedano, M. L., Ladero, V., Quevedo, C. et al. (2011). 'Role of tyramine synthesis by food-borne enterococcus durans in adaptation to the gastrointestinal tract environment'. *Applied and Environmental Microbiology* 77.2, pp. 699–702.
- De Preter, V., Machiels, K., Joossens, M., Arijis, I., Matthys, C. et al. (2015). 'Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD'. *Gut* 64.3, pp. 447–458.
- Deblon, N., Veyrat-Durebex, C., Bourgoin, L., Caillon, A., Bussier, A.-L. et al. (2011). 'Mechanisms of the anti-obesity effects of oxytocin in diet-induced obese rats'. *PloS one* 6.9, e25565.
- DeBosch, B. J., Kluth, O., Fujiwara, H., Schürmann, A. and Moley, K. (2014). 'Early-onset metabolic syndrome in mice lacking the intestinal uric acid transporter SLC2A9'. *Nature Communications* 5, p. 4642.
- Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C. et al. (2011). 'Evaluation of the Infinium Methylation 450K technology'. *Epigenomics* 3.6, pp. 771–784.
- Deelen, J., Beekman, M., Uh, H. W., Helmer, Q., Kuningas, M. et al. (2011). 'Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited'. *Aging Cell* 10.4, pp. 686–698.
- Deguchi, T., Ohtsuki, S., Otagiri, M., Takanaga, H., Asaba, H. et al. (2002). 'Major role of organic anion transporter 3 in the transport of indoxyl sulfate in the kidney'. *Kidney International* 61.5, pp. 1760–1768.
- Dehaven, C. D., Evans, A. M., Dai, H. and Lawton, K. A. (2010). 'Organization of GC/MS and LC/MS metabolomics data into chemical libraries'. *Journal of cheminformatics* 2.1, p. 9.
- Dennis, E. (1994). 'Diversity of group types, regulation, and function of phospholipase A2'. *Journal of Biological Chemistry* 269.18, pp. 13057–13060.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L. et al. (2006). 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB'. *Applied and Environmental Microbiology* 72.7, pp. 5069–5072.
- Després, J.-P. (2007). 'Cardiovascular disease under the influence of excess visceral fat'. *Critical pathways in cardiology* 6.2, pp. 51–9.
- Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H. (2006). 'Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics'. *Analytical chemistry* 78.13, pp. 4281–90.
- Do, K. T., Kastenmüller, G., Mook-Kanamori, D. O., Yousri, N. A., Theis, F. J. et al. (2015). 'Network-based approach for analyzing intra- and interfluid metabolite associations in human blood, urine, and saliva'. *Journal of proteome research* 14.2, pp. 1183–94.
- Doogue, M. P. and Polasek, T. M. (2011). 'Drug dosing in renal disease'. *The Clinical biochemist. Reviews* 32.2, pp. 69–73.
- Dreisbach, A. W. and Lertora, J. J. L. (2008). 'The effect of chronic renal failure on drug metabolism and transport'. *Expert opinion on drug metabolism & toxicology* 4.8, pp. 1065–74.
- Dumas, M. E., Maibaum, E. C., Teague, C., Ueshima, H., Zhou, B. et al. (2006). 'Assessment of analytical reproducibility of 1H NMR spectroscopy based metabolomics for large-scale epidemiological research: The INTERMAP study'. *Analytical Chemistry* 78.7, pp. 2199–2208.
- Duranton, F., Lundin, U., Gayraud, N., Mischak, H., Aparicio, M. et al. (2014). 'Plasma and urinary amino acid metabolomic profiling in patients with different levels of kidney function'. *Clinical Journal of the American Society of Nephrology* 9.1, pp. 37–45.
- Eddy, S. R. (2001). 'Non-coding RNA genes and the modern RNA world'. *Nature reviews. Genetics* 2.12, pp. 919–29.
- Edgar, R. C. (2010). 'Search and clustering orders of magnitude faster than BLAST'. *Bioinformatics* 26.19, pp. 2460–2461.
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. and Knight, R. (2011). 'UCHIME improves sensitivity and speed of chimera detection'. *Bioinformatics* 27.16, pp. 2194–2200.
- Ehrlich, M. (2002). 'DNA methylation in cancer: too much, but also too little'. *Oncogene* 21.35, pp. 5400–13.
- Ercan, A., Cui, J., Chatterton, D. E. W., Deane, K. D., Hazen, M. M. et al. (2010). 'Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis'. *Arthritis & Rheumatism* 62.8, pp. 2239–2248.
- Esser, N., Legrand-Poels, S., Piette, J., Scheen, A. J. and Paquot, N. (2014). 'Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes'. *Diabetes research and clinical practice* 105.2, pp. 141–50.
- Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. and Milgram, E. (2009). 'Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems'. *Analytical Chemistry* 81.16, pp. 6656–6667.

- Eyk, H. G., Van, Vermaat, R. J., Leijnse-Ybema, H. J. and Leijnse, B. (1968). 'The conversion of creatinine by creatininase of bacterial origin.' *Enzymologia* 34.3, pp. 198–202.
- Faisal, F. E. and Milenkovi, T. (2014). 'Dynamic networks reveal key players in aging'. *Bioinformatics* 30.12, pp. 1721–1729.
- Falconer, D. S. (1960). *Introduction to quantitative genetics*. New York: Ronald Press Co, p. 386.
- Farooqui, A. A. and Horrocks, L. A. (2007). *Glycerophospholipids in the Brain*. New York, NY: Springer New York.
- Fellinghauer, B., Bühlmann, P., Ryffel, M., Rhein, M. von and Reinhardt, J. D. (2013). 'Stable graphical model estimation with Random Forests for discrete, continuous, and mixed variables'. *Computational Statistics & Data Analysis* 64, pp. 132–152.
- Fernstrom, J. D. and Fernstrom, M. H. (2007). 'Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain.' *The Journal of nutrition* 137.6 Suppl 1, 1539S–1547S, discussion 1548S.
- Ferrara, C., Brünker, P., Suter, T., Moser, S., Püntener, U. et al. (2006a). 'Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β 1, 4-N-acetylglucosaminyltransferase III and Golgi α -mannosidase II'. *Biotechnology and Bioengineering* 93.5, pp. 851–861.
- Ferrara, C., Stuart, F., Sondermann, P., Brünker, P. and Umaña, P. (2006b). 'The carbohydrate at Fc γ maRIIIa Asn-162: An element required for high affinity binding to non-fucosylated IgG glycoforms'. *Journal of Biological Chemistry* 281.8, pp. 5032–5036.
- Fitzpatrick, A. L., Kronmal, R. a., Gardner, J. P., Psaty, B. M., Jenny, N. S. et al. (2007). 'Leukocyte telomere length and cardiovascular disease in the cardiovascular health study'. *American Journal of Epidemiology* 165.1, pp. 14–21.
- Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D. et al. (2013). 'Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach'. *Diabetes* 62.2, pp. 639–648.
- Fogarty, A. W., Glancy, C., Jones, S., Lewis, S. A., McKeever, T. M. et al. (2008). 'A prospective study of weight change and systemic inflammation over 9 y'. *The American journal of clinical nutrition* 87.1, pp. 30–5.
- Fortunato, S. (2010). 'Community detection in graphs'. *Physics Reports* 486.3–5, pp. 75–174.
- Fouque, D., Wang, P., Laville, M. and Boissel, J. P. (2001). 'Low protein diets for chronic renal failure in non diabetic adults.' *Cochrane database of systematic reviews (Online)* 2, p. CD001892.
- Fournier, N., Paul, J. L., Atger, V., Cogny, A., Soni, T. et al. (1997). 'HDL phospholipid content and composition as a major factor determining cholesterol efflux capacity from Fu5AH cells to human serum.' *Arteriosclerosis, thrombosis, and vascular biology* 17.11, pp. 2685–91.
- Fox, C. S., Massaro, J. M., Hoffmann, U., Pou, K. M., Maurovich-Horvat, P. et al. (2007). 'Abdominal visceral and subcutaneous adipose tissue compartments: Association with metabolic risk factors in the framingham heart study'. *Circulation* 116.1, pp. 39–48.
- Fox, C. J., Hammerman, P. S. and Thompson, C. B. (2005). 'Fuel feeds function: energy metabolism and the T-cell response.' *Nature reviews. Immunology* 5.11, pp. 844–52.
- Franceschi, C. and Campisi, J. (2014). 'Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated Diseases'. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 69.Suppl 1, S4–S9.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M. et al. (2013). 'STRING v9.1: protein-protein interaction networks, with increased coverage and integration.' *Nucleic acids research* 41.Database issue, pp. D808–15.
- Freddolino, P. L. and Tavazoie, S. (2012). 'The dawn of virtual cell biology.' *Cell* 150.2, pp. 248–50.
- Freitas, A. A. and Magalhães, J. P. de (2011). 'A review and appraisal of the DNA damage theory of ageing'. *Mutation Research/Reviews in Mutation Research* 728.1–2, pp. 12–22.
- Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C. et al. (2008). 'Microbial community gene expression in ocean surface waters.' *Proceedings of the National Academy of Sciences of the United States of America* 105.10, pp. 3805–10.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008). 'Sparse inverse covariance estimation with the graphical lasso.' *Biostatistics (Oxford, England)* 9.3, pp. 432–41.
- Friedman, N., Linial, M., Nachman, I. and Pe'er, D. (2000). 'Using Bayesian networks to analyze expression data.' *Journal of computational biology : a journal of computational molecular cell biology* 7, pp. 601–620.
- Fuente, A. de la, Bing, N., Hoeschele, I. and Mendes, P. (2004). 'Discovery of meaningful associations in genomic data using partial correlation coefficients'. *Bioinformatics* 20.18, pp. 3565–3574.
- Fuster, M. M. and Esko, J. D. (2005). 'The sweet and sour of cancer: glycans as novel therapeutic targets.' *Nature reviews. Cancer* 5.7, pp. 526–42.

- Garber, A. J. (1978). 'The regulation of skeletal muscle alanine and glutamine formation and release in experimental chronic uremia in the rat: subsensitivity of adenylate cyclase and amino acid release to epinephrine and serotonin.' *The Journal of clinical investigation* 62.3, pp. 633–41.
- Gatz, M., Reynolds, C. a., Fratiglioni, L., Johansson, B., Mortimer, J. a. et al. (2006). 'Role of genes and environments for explaining Alzheimer disease.' *Archives of general psychiatry* 63.2, pp. 168–74.
- Gelfi, C., Vigano, A., Ripamonti, M., Pontoglio, A., Begum, S. et al. (2006). 'The human muscle proteome in aging.' *Journal of proteome research* 5.6, pp. 1344–53.
- Giacomello, A., Di Sciascio, N. and Quarantino, C. P. (1997). 'Relation between serum triglyceride level, serum urate concentration, and fractional urate excretion.' *Metabolism: clinical and experimental* 46.9, pp. 1085–9.
- Glass, D., Viñuela, A., Davies, M. N., Ramasamy, A., Parts, L. et al. (2013). 'Gene expression changes with age in skin, adipose tissue, blood and brain.' *Genome biology* 14.7, R75.
- Goh, K.-i., Cusick, M. E., Valle, D., Childs, B., Vidal, M. et al. (2007). 'The human disease network.' *Proceedings of the National Academy of Sciences of the United States of America* 104.21, pp. 8685–90.
- Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R. et al. (2016). 'Genetic Determinants of the Gut Microbiome in UK Twins.' *Cell Host and Microbe* 19.5, pp. 731–743.
- Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O. et al. (2014). 'Human Genetics Shape the Gut Microbiome.' *Cell* 159.4, pp. 789–799.
- Gornik, O. and Lauc, G. (2008). 'Glycosylation of serum proteins in inflammatory diseases.' *Disease markers* 25.4-5, pp. 267–278.
- Gornik, O., Pavić, T. and Lauc, G. (2012). 'Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease.' *Biochimica et biophysica acta* 1820.9, pp. 1318–26.
- Gornik, O., Wagner, J., Pucić, M., Knezević, A., Redžić, I. et al. (2009). 'Stability of N-glycan profiles in human plasma.' *Glycobiology* 19.12, pp. 1547–53.
- Grabe, H. J., Assel, H., Bahls, T., Dörr, M., Endlich, K. et al. (2014). 'Cohort profile: Greifswald approach to individualized medicine (GANI_MED).' *Journal of translational medicine* 12.1, p. 144.
- Gregg, E. W., Li, Y., Wang, J., Burrows, N. R., Ali, M. K. et al. (2014). 'Changes in diabetes-related complications in the United States, 1990-2010'. *N Engl J Med* 370.16, pp. 1514–1523.
- Grice, E. a. and Segre, J. a. (2012). 'The human microbiome: our second genome.' *Annual review of genomics and human genetics* 13, pp. 151–70.
- Grundberg, E., Meduri, E., Sandling, J. K., Hedman, A. K., Keildson, S. et al. (2013). 'Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements.' *American journal of human genetics* 93.5, pp. 876–90.
- Grundberg, E., Small, K. S., Hedman, Å. K., Nica, A. C., Buil, A. et al. (2012). 'Mapping cis- and trans-regulatory effects across multiple tissues in twins.' *Nature genetics* 44.10, pp. 1084–9.
- Guo, J., Levina, E., Michailidis, G. and Zhu, J. (2010). 'Joint Structure Estimation for Categorical Markov Networks'.
- Guolo, A. and Varin, C. (2015). 'Random-effects meta-analysis: the number of studies matters'. *Statistical Methods in Medical Research*, p. 096228021558356.
- Gustafsson, M., Nestor, C. E., Zhang, H., Barabási, A.-L., Baranzini, S. et al. (2014). 'Modules, networks and systems medicine for understanding disease and aiding diagnosis.' *Genome medicine* 6.10, p. 82.
- Ha, N.-T., Freytag, S. and Bickeboeller, H. (2014). 'Coverage and efficiency in current SNP chips'. *European Journal of Human Genetics* 22.9, pp. 1124–1130.
- Haarbo, J., Gotfredsen, A., Hassager, C. and Christiansen, C. (1991). 'Validation of body composition by dual energy X-ray absorptiometry (DEXA)'. *Clinical Physiology* 11.4, pp. 331–341.
- Hammond, C. J., Duncan, D. D., Snieder, H., Lange, M. de, West, S. K. et al. (2001). 'The heritability of age-related cortical cataract: the twin eye study.' *Investigative ophthalmology & visual science* 42.3, pp. 601–5.
- Hanhineva, K., Lankinen, M. a., Pedret, A., Schwab, U., Kolehmainen, M. et al. (2015). 'Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, fatty fish, and bilberries in a randomized controlled trial.' *The Journal of nutrition* 145.1, pp. 7–17.
- Harbo, M., Bendix, L., Bay-Jensen, A.-C., Graakjaer, J., Sørensen, K. et al. (2012). 'The distribution pattern of critically short telomeres in human osteoarthritic knees'. *Arthritis Research & Therapy* 14.1, R12.
- Harley, C. B., Vaziri, H., Counter, C. M. and Allsopp, R. C. (1992). 'The telomere hypothesis of cellular aging.' *Experimental gerontology* 27.4, pp. 375–82.
- Harley, C. B., Futcher, A. B. and Greider, C. W. (1990). 'Telomeres shorten during ageing of human fibroblasts.' *Nature* 345.6274, pp. 458–60.
- Harman, D. (1956). 'Aging: A Theory Based on Free Radical and Radiation Chemistry'. *Journal of Gerontology* 11.3, pp. 298–300.
- Harman, D. (2001). 'Aging: overview.' *Annals of the New York Academy of Sciences* 928, pp. 1–21.
- (1988). 'The aging process.' *Basic life sciences* 49.11, pp. 1057–65.

- Harst, P. van der, Wong, L. S., Boer, R. A. de, Brouillette, S. W., Steege, G. van der et al. (2008). 'Possible Association Between Telomere Length and Renal Dysfunction in Patients With Chronic Heart Failure'. *The American Journal of Cardiology* 102.2, pp. 207–210.
- Hauner, H. (2004). 'The new concept of adipose tissue function.' *Physiology & behavior* 83.4, pp. 653–8.
- Hoek, J. B. and Taraschi, T. F. (1988). 'Cellular adaptation to ethanol.' *Trends in biochemical sciences* 13.7, pp. 269–74.
- Höfling, H. and Tibshirani, R. (2009). 'Estimation of Sparse Binary Pairwise Markov Networks using Pseudo-likelihoods.' *Journal of machine learning research : JMLR* 10, pp. 883–906.
- Hojs, R., Bevc, S., Ekart, R., Gorenjak, M. and Puklavec, L. (2006). 'Serum cystatin C as an endogenous marker of renal function in patients with mild to moderate impairment of kidney function'. *Nephrology Dialysis Transplantation* 21.7, pp. 1855–1862.
- Holle, R., Happich, M., Löwel, H. and Wichmann, H. (2005). 'KORA - A Research Platform for Population Based Health Research'. *Gesundheitswesen* 67.S 01, pp. 19–25.
- Horvath, S. (2013). 'DNA methylation age of human tissues and cell types.' *Genome biology* 14.10, R115.
- Houle, D., Govindaraju, D. R. and Omholt, S. (2010). 'Phenomics: the next challenge.' *Nature reviews. Genetics* 11.12, pp. 855–66.
- Hsu, C.-y., McCulloch, C. E., Iribarren, C., Darbinian, J. and Go, A. S. (2006). 'Annals of Internal Medicine Article Body Mass Index and Risk for End-Stage Renal Disease'. *Annals of Internal Medicine* 144.1, pp. 21–28.
- Hu, X.-X., Zhong, L., Zhang, X., Gao, Y.-M. and Liu, B.-Z. (2014). 'NLS-RARalpha Promotes Proliferation and Inhibits Differentiation in HL-60 Cells'. *International Journal of Medical Sciences* 11.3, pp. 247–254.
- Huang, H. and Bader, J. S. (2009). 'Precision and recall estimates for two-hybrid screens.' *Bioinformatics (Oxford, England)* 25.3, pp. 372–8.
- Hubert, L. and Arabie, P. (1985). 'Comparing partitions'. *Journal of Classification* 2.1, pp. 193–218.
- Huffman, J. E., Pu i -Bakovi, M., Klari, L., Hennig, R., Selman, M. H. J. et al. (2014). 'Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research'. *Molecular & Cellular Proteomics* 13.6, pp. 1598–1610.
- Hulbert, A. (2003). 'Life, death and membrane bilayers.' *The Journal of experimental biology* 206.Pt 14, pp. 2303–2311.
- Humphries, M. D. and Gurney, K. (2008). 'Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence'. *PLoS ONE* 3.4. Ed. by Sporns, O., e0002051.
- Hunter, R. W., Ivy, J. R. and Bailey, M. a. (2014). 'Glucocorticoids and renal Na⁺ transport: implications for hypertension and salt sensitivity.' *The Journal of physiology* 592.Pt 8, pp. 1731–44.
- Ideker, T., Galitski, T. and Hood, L. (2001). 'A new approach to decoding life: systems biology.' *Annual review of genomics and human genetics* 2, pp. 343–72.
- Ideker, T., Ozier, O., Schwikowski, B. and Siegel, A. F. (2002). 'Discovering regulatory and signalling circuits in molecular interaction networks.' *Bioinformatics (Oxford, England)* 18 Suppl 1, S233–40.
- Imig, J. D. and Ryan, M. J. (2013). 'Immune and inflammatory role in renal disease.' *Comprehensive Physiology* 3.2, pp. 957–76.
- Ishimori, M. L., Altman, R. D., Cohen, M. J., Cui, J., Guo, X. et al. (2010). 'Heritability patterns in hand osteoarthritis: the role of osteophytes.' *Arthritis research & therapy* 12.5, R180.
- Issaq, H. J., Nativ, O., Waybright, T., Luke, B., Veenstra, T. D. et al. (2008). 'Detection of bladder cancer in human urine by metabolomic profiling using high performance liquid chromatography/mass spectrometry.' *The Journal of urology* 179.6, pp. 2422–6.
- Jackson, M., Jeffery, I. B., Beaumont, M., Bell, J. T., Clark, A. G. et al. (2016a). 'Signatures of early frailty in the gut microbiota.' *Genome medicine* 8.1, p. 8.
- Jackson, M. A., Bell, J. T., Spector, T. D. and Steves, C. J. (2016b). 'A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units'. *PeerJ* 4, e2341.
- Jaenisch, R. and Bird, A. (2003). 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.' *Nature genetics* 33 Suppl.march, pp. 245–54.
- Janda, J. M. and Abbott, S. L. (2007). '16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls.' *Journal of clinical microbiology* 45.9, pp. 2761–4.
- Jang, H. R. and Rabb, H. (2015). 'Immune cells in experimental acute kidney injury.' *Nature reviews. Nephrology* 11.2, pp. 88–101.
- Jeong, H., Mason, S. P., Barabási, a. L. and Oltvai, Z. N. (2001). 'Lethality and centrality in protein networks.' *Nature* 411.6833, pp. 41–2.
- Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S. et al. (2013). 'Chronic kidney disease: Global dimension and perspectives'. *The Lancet* 382.9888, pp. 260–272.
- Johansen, K. L., Mulligan, K. and Schambelan, M. (1999). 'Anabolic effects of nandrolone decanoate in patients receiving dialysis: a randomized controlled trial.' *JAMA* 281.14, pp. 1275–81.

- Johnson, R. J., Nakagawa, T., Jalal, D., Sanchez-Lozada, L. G., Kang, D.-H. et al. (2013). 'Uric acid and chronic kidney disease: which is chasing which?' *Nephrology Dialysis Transplantation* 28.9, pp. 2221–2228.
- Johnson, T. E. (2006). 'Recent results: biomarkers of aging.' *Experimental gerontology* 41.12, pp. 1243–6.
- Johnson, W. E., Li, C. and Rabinovic, A. (2007). 'Adjusting batch effects in microarray expression data using empirical Bayes methods.' *Biostatistics (Oxford, England)* 8.1, pp. 118–27.
- Jones, J. D. and Burnett, P. C. (1974). 'Creatinine metabolism in humans with decreased renal function: creatinine deficit.' *Clinical chemistry* 20.9, pp. 1204–12.
- Jovel, J. (2016). 'Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics'. *Evolutionary and Genomic Microbiology* 7. April, p. 459.
- Joy, M. P., Brock, A., Ingber, D. E. and Huang, S. (2005). 'High-Betweenness Proteins in the Yeast Protein Interaction Network'. *Journal of Biomedicine and Biotechnology* 2005.2, pp. 96–103.
- Kamada, T. and Otsuji, S. (1983). 'Lower levels of erythrocyte membrane fluidity in diabetic patients. A spin label study.' *Diabetes* 32.7, pp. 585–91.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012). 'KEGG for integration and interpretation of large-scale molecular data sets'. *eng. Nucleic acids research* 40.Database issue, pp. D109–114.
- Kaneko, Y., Nimmerjahn, F. and Ravetch, J. V. (2006). 'Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation.' *Science* 313.5787, pp. 670–3.
- Karmen, A., Wroblewski, F. and Ladue, J. S. (1955). 'Transaminase activity in human blood.' *The Journal of clinical investigation* 34.1, pp. 126–31.
- Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M. et al. (2012). 'A whole-cell computational model predicts phenotype from genotype.' *Cell* 150.2, pp. 389–401.
- Karsten, C. M., Pandey, M. K., Figge, J., Kilchenstein, R., Taylor, P. R. et al. (2012). 'Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of Fc γ RIIB and dectin-1'. *Nature Medicine* 18.9, pp. 1401–1406.
- Kaul, S., Rothney, M. P., Peters, D. M., Wacker, W. K., Davis, C. E. et al. (2012). 'Dual-energy X-ray absorptiometry for quantification of visceral fat.' *Obesity (Silver Spring, Md.)* 20.6, pp. 1313–8.
- Kerr, M., Bray, B., Medcalf, J., O'Donoghue, D. J. and Matthews, B. (2012). 'Estimating the financial cost of chronic kidney disease to the NHS in England'. *Nephrology Dialysis Transplantation* 27.Suppl 3, pp. iii73–80.
- Keser, T., Vučković, F., Barrios, C., Zierer, J., Wahl, A. et al. (2017b). 'Effects of statins on the immunoglobulin G glycome.' *Biochimica et biophysica acta* 1861.5 Pt A, pp. 1152–1158.
- Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S. et al. (2009). 'Human Protein Reference Database–2009 update.' *Nucleic acids research* 37.Database issue, pp. D767–72.
- Kim, I. H., Kisseleva, T. and Brenner, D. A. (2015). 'Aging and liver disease.' *Current opinion in gastroenterology* 31.3, pp. 184–91.
- Kim, S.-Y. and Volsky, D. J. (2005). 'PAGE: parametric analysis of gene set enrichment.' *BMC bioinformatics* 6.1, p. 144.
- Kimura, M., Hjelmberg, J. V. B., Gardner, J. P., Bathum, L., Brimacombe, M. et al. (2008). 'Telomere length and mortality: A study of leukocytes in elderly danish twins'. *American Journal of Epidemiology* 167.7, pp. 799–806.
- Kind, T., Cho, E., Park, T. D., Deng, N., Liu, Z. et al. (2016). 'Interstitial Cystitis-Associated Urinary Metabolites Identified by Mass-Spectrometry Based Metabolomics Analysis.' *Scientific reports* 6.1, p. 39227.
- Kirkwood, T. B. and Austad, S. N. (2000). 'Why do we age?' *Nature* 408.6809, pp. 233–8.
- Knezević, A., Polasek, O., Gornik, O., Rudan, I., Campbell, H. et al. (2009). 'Variability, heritability and environmental determinants of human plasma N-glycome.' *Journal of proteome research* 8.2, pp. 694–701.
- Kon, V., Yang, H. and Fazio, S. (2015). 'Residual Cardiovascular Risk in Chronic Kidney Disease: Role of High-density Lipoprotein.' *Archives of medical research* 46.5, pp. 379–91.
- Kong, H. H. (2011). 'Skin microbiome: genomics-based insights into the diversity and role of skin microbes.' *Trends in molecular medicine* 17.6, pp. 320–8.
- Kopple, J. D. (2007). 'Phenylalanine and tyrosine metabolism in chronic kidney failure.' *The Journal of nutrition* 137.6 Suppl 1, 1586S–1590S, discussion 1597S–1598S.
- Koropatkin, N. M., Cameron, E. a. and Martens, E. C. (2012). 'How glycan metabolism shapes the human gut microbiota.' *Nature reviews. Microbiology* 10.5, pp. 323–35.
- Kostic, A. D., Xavier, R. J. and Gevers, D. (2014). 'The microbiome in inflammatory bowel disease: current status and the future ahead.' *Gastroenterology* 146.6, pp. 1489–99.
- Köttgen, A., Pattaro, C., Böger, C. a., Fuchsberger, C., Olden, M. et al. (2010). 'New loci associated with kidney function and chronic kidney disease.' *Nature genetics* 42.5, pp. 376–384.
- Krakoff, L. R. (1988). 'Glucocorticoid excess syndromes causing hypertension.' *Cardiology clinics* 6.4, pp. 537–45.

- Kriete, A., Lechner, M., Clearfield, D. and Bohmann, D. (2011). 'Computational systems biology of aging.' *Wiley interdisciplinary reviews. Systems biology and medicine* 3.4, pp. 414–28.
- Kristic, J., Vuckovic, F., Menni, C., Klaric, L., Keser, T. et al. (2014). 'Glycans Are a Novel Biomarker of Chronological and Biological Ages'. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 69.7, pp. 779–789.
- Krumsiek, J., Mittelstrass, K., Do, K. T., Stücker, F., Ried, J. et al. (2015). 'Gender-specific pathway differences in the human serum metabolome.' *Metabolomics : Official journal of the Metabolomic Society* 11.6, pp. 1815–1833.
- Krumsiek, J., Suhre, K., Illig, T., Adamski, J. and Theis, F. J. (2011). 'Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data.' *BMC systems biology* 5.1, p. 21.
- Kuk, J. L., Katzmarzyk, P. T., Nichaman, M. Z., Church, T. S., Blair, S. N. et al. (2006). 'Visceral fat is an independent predictor of all-cause mortality in men.' *Obesity* 14.2, pp. 336–341.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C. et al. (2001). 'Initial sequencing and analysis of the human genome'. *Nature* 409.6822, pp. 860–921.
- Langefeld, C. D., Beck, S. R., Bowden, D. W., Rich, S. S., Wagenknecht, L. E. et al. (2004). 'Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus.' *American journal of kidney diseases : the official journal of the National Kidney Foundation* 43.5, pp. 796–800.
- Langfelder, P. and Horvath, S. (2008). 'WGCNA: an R package for weighted correlation network analysis'. *BMC Bioinformatics* 9.1, p. 559.
- Lasisi, T. J., Raji, Y. R. and Salako, B. L. (2016). 'Salivary creatinine and urea analysis in patients with chronic kidney disease: a case control study.' *BMC nephrology* 17.10, p. 10.
- Lauc, G., Essafi, A., Huffman, J. E., Hayward, C., Knežević, A. et al. (2010). 'Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1 α as a master regulator of plasma protein fucosylation.' *PLoS genetics* 6.12, e1001256.
- Lauc, G., Huffman, J. E., Pučić, M., Zgaga, L., Adamczyk, B. et al. (2013). 'Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers.' *PLoS genetics* 9.1, e1003225.
- Lauritzen, S. L. and Wermuth, N. (1989). 'Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative'. *The Annals of Statistics* 17.1, pp. 31–57.
- Lawson, E. a., Marengi, D. a., DeSanti, R. L., Holmes, T. M., Schoenfeld, D. a. et al. (2015). 'Oxytocin reduces caloric intake in men.' *Obesity (Silver Spring, Md.)* 23.5, pp. 950–6.
- Lawton, K. A., Berger, A., Mitchell, M., Milgram, K. E., Evans, A. M. et al. (2008). 'Analysis of the adult human plasma metabolome.' *Pharmacogenomics* 9.4, pp. 383–97.
- Lea, J. P. and Nicholas, S. B. (2002). 'Diabetes mellitus and hypertension: key risk factors for kidney disease.' *Journal of the National Medical Association* 94.8 Suppl, 7S–15S.
- Leblond, F. A., Giroux, L., Villeneuve, J. P. and Pichette, V. (2000). 'Decreased in vivo metabolism of drugs in chronic renal failure.' *Drug metabolism and disposition: the biological fate of chemicals* 28.11, pp. 1317–20.
- Leblond, F., Guévin, C., Demers, C., Pellerin, I., Gascon-Barré, M. et al. (2001). 'Downregulation of hepatic cytochrome P450 in chronic renal failure.' *Journal of the American Society of Nephrology : JASN* 12.2, pp. 326–32.
- Lecocq, F. R., Mebane, D. and Madison, L. L. (1964). 'The Acute Effect of Hydrocortisone on Hepatic Glucose Output and Peripheral Glucose Utilization'. *Journal of Clinical Investigation* 43.2, pp. 237–246.
- Ledoit, O. and Wolf, M. (2003). 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection'. *Journal of Empirical Finance* 10.5, pp. 603–621.
- Lee, J. D. and Hastie, T. J. (2015). 'Learning the Structure of Mixed Graphical Models'. *Journal of Computational and Graphical Statistics* 24.1, pp. 230–253.
- Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N. et al. (1999). 'A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group.' *Annals of internal medicine* 130.6, pp. 461–70.
- Levey, A. S. and Coresh, J. (2012). 'Chronic kidney disease'. *The Lancet* 379.9811, pp. 165–180.
- Levey, A. S., Eckardt, K. U., Tsukamoto, Y., Levin, A., Coresh, J. et al. (2005). 'Definition and classification of chronic kidney disease: A position statement from Kidney Disease: Improving Global Outcomes (KDIGO)'. *Kidney International* 67.6, pp. 2089–2100.
- Levey, A. S., Stevens, L. a., Schmid, C. H., Zhang, Y. L., Castro, A. F. et al. (2009). 'A new equation to estimate glomerular filtration rate.' *Annals of internal medicine* 150.9, pp. 604–612.
- Levy, M. Z., Allsopp, R. C., Fletcher, A., Greider, C. W. and Harley, C. B. (1992). 'Telomere end-replication problem and cell aging'. *Journal of Molecular Biology* 225.4, pp. 951–960.
- Lewington, S., Whitlock, G., Clarke, R., Sherliker, P., Emberson, J. et al. (2007). 'Cholesterol and Vascular Mortality By Age, Sex, and Blood Pressure: a Meta-Analysis of Individual Data From 61 Prospective Studies With 55,000 Vascular'. *Lancet* 370.9602, pp. 1829–1839.

- Ley, R., Turnbaugh, P., Klein, S. and Gordon, J. (2006). 'Microbial ecology: human gut microbes associated with obesity.' *Nature* 444.7122, pp. 1022–3.
- Li, D., Morales, A., Gonzalez-Quintana, J., Norton, N., Siegfried, J. D. et al. (2010). 'Identification of novel mutations in RBM20 in patients with dilated cardiomyopathy.' *Clinical and translational science* 3.3, pp. 90–7.
- Li, H. (2011). 'A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data'. *Bioinformatics* 27.21, pp. 2987–2993.
- Li, J. and Ji, L. (2005). 'Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix.' *Heredity* 95.3, pp. 221–227.
- Li, M., Foo, J.-N., Wang, J.-Q., Low, H.-Q., Tang, X.-Q. et al. (2015). 'Identification of new susceptibility loci for IgA nephropathy in Han Chinese'. *Nature Communications* 6.1, p. 7270.
- Liu, H., Roeder, K. and Wasserman, L. (2010). 'Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models.' *Advances in neural information processing systems* 24.2, pp. 1432–1440.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y. et al. (2012). 'Comparison of Next-Generation Sequencing Systems'. *Journal of Biomedicine and Biotechnology* 2012, pp. 1–11.
- Livote, E. (2009). 'Introduction to Structural Equation Modeling Using SPSS and AMOS. Niels J. Blunch. Thousand Oaks, CA: Sage, 2008, 270 pages, \$39.95.' *Structural Equation Modeling: A Multidisciplinary Journal* 16.3, pp. 556–560.
- Long, T., Hicks, M., Yu, H.-C., Biggs, W. H., Kirkness, E. F. et al. (2017a). 'Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites.' *Nature genetics* 49.4, pp. 568–578.
- Longo, V. D., Mitteldorf, J. and Skulachev, V. P. (2005). 'Opinion: Programmed and altruistic ageing'. *Nature Reviews Genetics* 6.11, pp. 860–866.
- López-Otín, C., Blasco, M. a., Partridge, L., Serrano, M. and Kroemer, G. (2013). 'The hallmarks of aging.' *Cell* 153.6, pp. 1194–217.
- Lopez-Parra, V., Mallavia, B., Lopez-Franco, O., Ortiz-Munoz, G., Oguiza, A. et al. (2012). 'Fc Receptor Deficiency Attenuates Diabetic Nephropathy'. *Journal of the American Society of Nephrology* 23.9, pp. 1518–1527.
- Lorraine, M. and Cidlowski, J. (2003). 'Physiologic and Pharmacologic Effects of Corticosteroids'. *Cancer Medicine*. Ed. by Kufe, D. W., Pollock, R. E., Weichselbaum, R. R., Bast, R. C., Gansler, T. S. et al. 6th editio. Hamilton (ON): BC Decker.
- Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K. et al. (2012). 'Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010'. *The Lancet* 380.9859, pp. 2095–2128.
- Lozupone, C. and Knight, R. (2005). 'UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities'. *Applied and environmental microbiology* 71.12, pp. 8228–8235.
- Lu, T., Pan, Y., Kao, S.-y., Li, C., Kohane, I. et al. (2004). 'Gene regulation and DNA damage in the ageing human brain.' *Nature* 429.6994, pp. 883–91.
- Lv, J., Xu, D., Perkovic, V., Ma, X., Johnson, D. W. et al. (2012). 'Corticosteroid Therapy in IgA Nephropathy'. *Journal of the American Society of Nephrology* 23.6, pp. 1108–1116.
- Maeda, K., Mehta, H., Drevets, D. a. and Coggeshall, K. M. (2010). 'IL-6 increases B-cell IgG production in a feed-forward proinflammatory mechanism to skew hematopoiesis and elevate myeloid production.' *Blood* 115.23, pp. 4699–706.
- Magalhães, J. P. de, Curado, J. and Church, G. M. (2009). 'Meta-analysis of age-related gene expression profiles identifies common signatures of aging.' *Bioinformatics (Oxford, England)* 25.7, pp. 875–81.
- Magnúsdóttir, S., Ravcheev, D., Crécy-Lagard, V. de and Thiele, I. (2015). 'Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes.' *Frontiers in genetics* 6.MAR, p. 148.
- Magnusson, I., Rothman, D. L., Katz, L. D., Shulman, R. G. and Shulman, G. I. (1992). 'Increased rate of gluconeogenesis in type II diabetes mellitus. A ¹³C nuclear magnetic resonance study.' *The Journal of clinical investigation* 90.4, pp. 1323–7.
- Maher, P. (2005). 'The effects of stress and aging on glutathione metabolism'. *Ageing Research Reviews* 4.2, pp. 288–314.
- Mäkinen, V.-P., Tynkkynen, T., Soininen, P., Peltola, T., Kangas, A. J. et al. (2012). 'Metabolic diversity of progressive kidney disease in 325 patients with type 1 diabetes (the FinnDiane Study)'. *Journal of proteome research* 11.3, pp. 1782–90.
- Mäkinen, V.-P., Soininen, P., Kangas, A. J., Forsblom, C., Tolonen, N. et al. (2013). 'Triglyceride-cholesterol imbalance across lipoprotein subclasses predicts diabetic kidney disease and mortality in type 1 diabetes: the FinnDiane Study.' *Journal of internal medicine* 273.4, pp. 383–95.

- Malhotra, R., Wormald, M. R., Rudd, P. M., Fischer, P. B., Dwek, R. A. et al. (1995). 'Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein.' *Nature medicine* 1.3, pp. 237–43.
- Manini, P., De Palma, G., Andreoli, R., Goldoni, M., Poli, D. et al. (2003). 'Urinary excretion of 4-vinyl phenol after experimental and occupational exposure to styrene'. *Giornale italiano di medicina del lavoro ed ergonomia* 25 Suppl.3, pp. 61–62.
- Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J. et al. (2012). 'Wisdom of crowds for robust gene network inference.' *Nature methods* 9.8, pp. 796–804.
- Marcobal, A., Kashyap, P. C., Nelson, T. A., Aronov, P. A., Donia, M. S. et al. (2013). 'A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice.' *The ISME journal* 7.10, pp. 1933–43.
- Markley, J. L., Brüschweiler, R., Edison, A. S., Eghbalnia, H. R., Powers, R. et al. (2017). 'The future of NMR-based metabolomics'. *Current Opinion in Biotechnology* 43, pp. 34–40.
- Martin, N. G. and Eaves, L. J. (1977). 'The genetical analysis of covariance structure'. *Heredity* 38.1, pp. 79–95.
- Masuda, K., Kubota, T., Kaneko, E., Iida, S., Wakitani, M. et al. (2007). 'Enhanced binding affinity for FcγRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity'. *Molecular Immunology* 44.12, pp. 3122–3131.
- Masuo, K., Kawaguchi, H., Mikami, H., Ogihara, T. and Tuck, M. L. (2003). 'Serum uric acid and plasma norepinephrine concentrations predict subsequent weight gain and blood pressure elevation.' *Hypertension (Dallas, Tex. : 1979)* 42.4, pp. 474–80.
- Mather, K. A., Jorm, A. F., Parslow, R. A. and Christensen, H. (2011). 'Is Telomere Length a Biomarker of Aging? A Review'. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 66A.2, pp. 202–213.
- Matsumoto, M., Kibe, R., Ooga, T., Aiba, Y., Kurihara, S. et al. (2012). 'Impact of Intestinal Microbiota on Intestinal Luminal Metabolome'. *Scientific Reports* 2, p. 233.
- May, R. C., Kelly, R. A. and Mitch, W. E. (1987). 'Mechanisms for defects in muscle protein metabolism in rats with chronic uremia. Influence of metabolic acidosis'. *Journal of Clinical Investigation* 79.4, pp. 1099–1103.
- Mazumder, R. and Hastie, T. (2012). 'The graphical lasso: New insights and alternatives'. *Electronic Journal of Statistics* 6.August, pp. 2125–2149.
- McAuley, M. T., Kenny, R. A., Kirkwood, T. B. L., Wilkinson, D. J., Jones, J. J. L. et al. (2009). 'A mathematical model of aging-related and cortisol induced hippocampal dysfunction.' *BMC neuroscience* 10, p. 26.
- McClintock, M. K., Dale, W., Laumann, E. O. and Waite, L. (2016). 'Empirical redefinition of comprehensive health and well-being in the older adults of the United States'. *Proceedings of the National Academy of Sciences* 113.22, E3071–E3080.
- McHardy, I. H., Goudarzi, M., Tong, M., Ruegger, P. M., Schwager, E. et al. (2013). 'Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships.' *Microbiome* 1.1, p. 17.
- Mecocci, P., Cherubini, A., Beal, M. F., Cecchetti, R., Chionne, F. et al. (1996). 'Altered mitochondrial membrane fluidity in AD brain'. *Neuroscience Letters* 207.2, pp. 129–132.
- Mehra, V. C., Ramgolam, V. S. and Bender, J. R. (2005). 'Cytokines and cardiovascular disease.' *Journal of leukocyte biology* 78.4, pp. 805–18.
- Meinshausen, N. and Bühlmann, P. (2006). 'High-dimensional graphs and variable selection with the Lasso'. *The Annals of Statistics* 34.3, pp. 1436–1462.
- Meissner, A. (2010). 'Epigenetic modifications in pluripotent and differentiated cells.' *Nature biotechnology* 28.10, pp. 1079–88.
- Menni, C., Fauman, E., Erte, I., Perry, J. R. B., Kastenmüller, G. et al. (2013a). 'Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach'. *Diabetes* 62, pp. 4270–4276.
- Menni, C., Kastenmüller, G., Petersen, A. K., Bell, J. T., Psatha, M. et al. (2013b). 'Metabolomic markers reveal novel pathways of ageing and early development in human populations.' *International journal of epidemiology* 42.4, pp. 1111–9.
- Menni, C., Keser, T., Mangino, M., Bell, J. T., Erte, I. et al. (2013c). 'Glycosylation of immunoglobulin g: role of genetic and epigenetic influences.' *PloS one* 8.12, e82558.
- Menni, C., Kiddle, S. J., Mangino, M., Vinuela, A., Psatha, M. et al. (2015). 'Circulating Proteomic Signatures of Chronological Age'. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 70.7, pp. 809–816.
- Menni, C., Migaud, M., Glastonbury, C. A., Beaumont, M., Nikolaou, A. et al. (2016). 'Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health'. *Obesity* 24.6, pp. 1380–1388.
- Menni, C., Zierer, J., Valdes, A. M. and Spector, T. D. (2017b). 'Mixing omics: combining genetics and metabolomics to study rheumatic diseases'. *Nature Reviews Rheumatology* 13.3, pp. 174–181.

- Menon, V., Shlipak, M. G., Wang, X., Coresh, J., Greene, T. et al. (2007). 'Cystatin C as a risk factor for outcomes in chronic kidney disease.' *Annals of internal medicine* 147.1, pp. 19–27.
- Miller, J. a., Oldham, M. C. and Geschwind, D. H. (2008). 'A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging.' *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28.6, pp. 1410–20.
- Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R. et al. (2005). 'Standardisation of spirometry.' *The European respiratory journal* 26.2, pp. 319–38.
- Mischak, H., Delles, C., Vlahou, A. and Vanholder, R. (2015). 'Proteomic biomarkers in kidney disease: issues in development and implementation.' *Nature reviews. Nephrology* 11.4, pp. 221–32.
- Mitnitski, A. B., Mogilner, A. J. and Rockwood, K. (2001). 'Accumulation of Deficits as a Proxy Measure of Aging'. *The Scientific World JOURNAL* 1, pp. 323–336.
- Mitnitski, A., Song, X. and Rockwood, K. (2013). 'Assessing biological aging: the origin of deficit accumulation.' *Biogerontology* 14.6, pp. 709–17.
- Moayyeri, A., Hammond, C. J., Valdes, A. M. and Spector, T. D. (2013). 'Cohort Profile: TwinsUK and Healthy Ageing Twin Study'. *International Journal of Epidemiology* 42.1, pp. 76–85.
- Moco, S., Vervoort, J., Moco, S., Bino, R. J., De Vos, R. C. et al. (2007). 'Metabolomics technologies and metabolite identification'. *TrAC Trends in Analytical Chemistry* 26.9, pp. 855–866.
- Møller, N., Meek, S., Bigelow, M., Andrews, J. and Nair, K. S. (2000). 'The kidney is an important site for in vivo phenylalanine-to-tyrosine conversion in adult humans: A metabolic role of the kidney.' *Proceedings of the National Academy of Sciences of the United States of America* 97.3, pp. 1242–6.
- Moss, C. W., Lambert, M. a. and Goldsmith, D. J. (1970). 'Production of hydrocinnamic acid by clostridia.' *Applied microbiology* 19.2, pp. 375–8.
- Mueller, B., Klemm, E. J., Spooner, E., Claessen, J. H. and Ploegh, H. L. (2008). 'SEL1L nucleates a protein complex required for dislocation of misfolded glycoproteins.' *Proceedings of the National Academy of Sciences of the United States of America* 105.34, pp. 12325–30.
- Mungreiphy, N. K., Kapoor, S. and Sinha, R. (2011). 'Association between BMI, Blood Pressure, and Age: Study among Tangkhul Naga Tribal Males of Northeast India'. *Journal of Anthropology* 2011, pp. 1–6.
- Murabito, J. M., Yuan, R. and Lunetta, K. L. (2012). 'The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals.' *The journals of gerontology. Series A, Biological sciences and medical sciences* 67.5, pp. 470–9.
- Murata, K., Baumann, N. A., Saenger, A. K., Larson, T. S., Rule, A. D. et al. (2011). 'Relative performance of the MDRD and CKD-EPI equations for estimating glomerular filtration rate among patients with varied clinical presentations.' *Clinical journal of the American Society of Nephrology : CJASN* 6.8, pp. 1963–72.
- Nakanishi, N., Suzuki, K. and Tatara, K. (2003). 'Alcohol Consumption and Risk for Development of Impaired Fasting Glucose or Type 2 Diabetes in Middle-Aged Japanese Men'. *Diabetes Care* 26.1, pp. 48–54.
- Neale, M. C. and Miller, M. B. (1997). 'The use of likelihood-based confidence intervals in genetic models.' *Behavior genetics* 27.2, pp. 113–20.
- Ng Hublin, J. S. Y., Ryan, U., Trengove, R. and Maker, G. (2013). 'Metabolomic profiling of faecal extracts from *Cryptosporidium parvum* infection in experimental mouse models.' *PloS one* 8.10, e77803.
- Niewczas, M. A., Sirich, T. L., Mathew, A. V., Skupien, J., Mohny, R. P. et al. (2014). 'Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study'. *Kidney International* 85.5, pp. 1214–1224.
- Nihlén, U., Montn  mery, P., Lindholm, L. H. and L  fdahl, C. G. (2001). 'Increased serum levels of carbohydrate-deficient transferrin in patients with chronic obstructive pulmonary disease.' *Scandinavian journal of clinical and laboratory investigation* 61.5, pp. 341–7.
- Niwa, T. (1996). 'Organic acids and the uremic syndrome: protein metabolite hypothesis in the progression of chronic renal failure.' *Seminars in nephrology* 16.3, pp. 167–82.
- Novak, J., Julian, B. A., Mestecky, J. and Renfrow, M. B. (2012). 'Glycosylation of IgA1 and pathogenesis of IgA nephropathy'. *Seminars in Immunopathology* 34.3, pp. 365–382.
- Novak, J., Tomana, M., Matousov  , K., Brown, R., Hall, S. et al. (2005). 'IgA1-containing immune complexes in IgA nephropathy differentially affect proliferation of mesangial cells'. *Kidney International* 67.2, pp. 504–513.
- Novokmet, M., Luki  , E., Vu  kovi  , F.,   uri  ,   ., Keser, T. et al. (2014). 'Changes in IgG and total plasma protein glycomes in acute systemic inflammation'. *Scientific Reports* 4, pp. 1–10.
- Ny  ki, A., Buclin, T., Biollaz, J. and Decosterd, L. A. (2003). 'NAT2 and CYP1A2 phenotyping with caffeine: head-to-head comparison of AFMU vs. AAMU in the urine metabolite ratios.' *British journal of clinical pharmacology* 55.1, pp. 62–7.
- O'Callaghan, C. a., Shine, B. and Lasserson, D. S. (2011). 'Chronic kidney disease: a large-scale population-based study of the effects of introducing the CKD-EPI formula for eGFR reporting'. *BMJ Open* 1.2, e000308–e000308.

- Oeppen, J. and Vaupel, J. W. (2002). 'Demography. Broken limits to life expectancy.' *Science (New York, N.Y.)* 296.5570, pp. 1029–31.
- Office for National Statistics (2014). *Mortality in the United Kingdom, 1983–2013*. Tech. rep. Office for National Statistics.
- O'Hara, A. M. and Shanahan, F. (2006). 'The gut flora as a forgotten organ'. *EMBO Rep* 7.7, pp. 688–693.
- Ohnishi, A., Branch, R. A., Jackson, K., Hamilton, R., Biaggioni, I. et al. (1986). 'Chronic caffeine administration exacerbates renovascular, but not genetic, hypertension in rats'. *Journal of Clinical Investigation* 78.4, pp. 1045–1050.
- Okada, Y., Sim, X., Go, M. J., Wu, J.-Y., Gu, D. et al. (2012). 'Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations'. *Nature Genetics* 44.8, pp. 904–909.
- O'Keefe, S., Gzel, A., Drury, R., Cullina, M., Grealley, J. et al. (1991). 'Immunoglobulin G subclasses and spirometry in patients with chronic obstructive pulmonary disease.' *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 4.8, pp. 932–936.
- Olovnikov, A. M. (1996). 'Telomeres, telomerase, and aging: origin of the theory.' *Experimental gerontology* 31.4, pp. 443–8.
- Orentreich, N., Brind, J. L., Vogelmann, J. H., Andres, R. and Baldwin, H. (1992). 'Long-term longitudinal measurements of plasma dehydroepiandrosterone sulfate in normal men.' *The Journal of clinical endocrinology and metabolism* 75.4, pp. 1002–4.
- Orešič, M., Hyötyläinen, T., Herukka, S.-K., Sysi-Aho, M., Mattila, I. et al. (2011). 'Metabolome in progression to Alzheimer's disease.' *Translational psychiatry* 1.November, e57.
- Osswald, H. and Schnermann, J. (2011). 'Methylxanthines and the kidney.' *Methylxanthines, Handbook of experimental pharmacology*. Ed. by Fredholm, B. 269th ed. Vol. 200. Berlin Heidelberg: Springer, pp. 391–412.
- O'Toole, P. W. and Claesson, M. J. (2010). 'Gut microbiota: Changes throughout the lifespan from infancy to elderly'. *International Dairy Journal* 20.4, pp. 281–291.
- Ouweland, A. C., Bergsma, N., Parhiala, R., Lahtinen, S., Gueimonde, M. et al. (2008). 'Bifidobacterium microbiota and parameters of immune function in elderly subjects.' *FEMS immunology and medical microbiology* 53.1, pp. 18–25.
- Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I. et al. (2005). 'The MIPS mammalian protein-protein interaction database.' *Bioinformatics (Oxford, England)* 21.6, pp. 832–4.
- Pallet, N., Thervet, E., Beaune, P., Karras, A. and Bertho, G. (2014). 'The urinary metabolome of chronic kidney disease.' *Kidney international* 85.5, pp. 1239–40.
- Pandur, S., Pankiv, S., Johannessen, M., Moens, U. and Huseby, N.-E. (2007). 'Gamma-glutamyltransferase is upregulated after oxidative stress through the Ras signal transduction pathway in rat colon carcinoma cells'. *Free radical research* 41.12, pp. 1376–1384.
- Panossian, L. (2003). 'Telomere shortening in T cells correlates with Alzheimer's disease status'. *Neurobiology of Aging* 24.1, pp. 77–84.
- Parekh, P. J., Balart, L. a. and Johnson, D. a. (2015). 'The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease'. *Clinical and Translational Gastroenterology* 6.6, e91.
- Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A. et al. (1985). 'Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG.' *Nature* 316.6027, pp. 452–7.
- Pattaro, C., Köttgen, A., Teumer, A., Garnaas, M., Böger, C. A. et al. (2012). 'Genome-wide association and functional follow-up reveals new loci for kidney function.' *PLoS genetics* 8.3, e1002584.
- Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M. et al. (2016). 'Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function.' *Nature communications* 7, p. 10023.
- Payne, A. N., Chassard, C. and Lacroix, C. (2012). 'Gut microbial adaptation to dietary consumption of fructose, artificial sweeteners and sugar alcohols: implications for host-microbe interactions contributing to obesity.' *Obesity reviews : an official journal of the International Association for the Study of Obesity* 13.9, pp. 799–809.
- Peckett, A. J., Wright, D. C. and Riddell, M. C. (2011). 'The effects of glucocorticoids on adipose tissue lipid metabolism.' *Metabolism: clinical and experimental* 60.11, pp. 1500–10.
- Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyötyläinen, T., Nielsen, T. et al. (2016). 'Human gut microbes impact host serum metabolome and insulin sensitivity.' *Nature* 535.7612, pp. 376–81.
- Perrone, R. D., Madias, N. E. and Levey, A. S. (1992). 'Serum creatinine as an index of renal function: New insights into old concepts'. *Clinical Chemistry* 38.10, pp. 1933–1953.
- Petersen, a.-K. A. K., Zeilinger, S., Kastenmüller, G., Werner, R. M., Brügger, M. et al. (2014). 'Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits'. *Human Molecular Genetics* 23.2, pp. 534–545.

- Petersen, A.-K., Krumsiek, J., Wägele, B., Theis, F. J., Wichmann, H.-E. et al. (2012). 'On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies.' *BMC bioinformatics* 13.1, p. 120.
- Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L. et al. (2009). 'The NIH Human Microbiome Project.' *Genome research* 19.12, pp. 2317–23.
- Pineda Torra, I., Gervois, P. and Staels, B. (1999). 'Peroxisome proliferator-activated receptor alpha in metabolic disease, inflammation, atherosclerosis and aging.' *Current opinion in lipidology* 10.2, pp. 151–9.
- Pitts, R. F., Damian, A. C. and MacLeod, M. B. (1970). 'Synthesis of serine by rat kidney in vivo and in vitro.' *The American journal of physiology* 219.3, pp. 584–9.
- Portela, A. and Esteller, M. (2010). 'Epigenetic modifications and human disease.' *Nature biotechnology* 28.10, pp. 1057–68.
- Posada-Ayala, M., Zubiri, I., Martin-Lorenzo, M., Sanz-Maroto, A., Molero, D. et al. (2014). 'Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease.' *Kidney international* 85.1, pp. 103–11.
- Poynter, M. E. and Daynes, R. A. (1998). 'Peroxisome proliferator-activated receptor alpha activation modulates cellular redox status, represses nuclear factor-kappaB signaling, and reduces inflammatory cytokine production in aging.' *The Journal of biological chemistry* 273.49, pp. 32833–41.
- Prakash, S. and O'Hare, A. M. (2009). 'Interaction of Aging and Chronic Kidney Disease'. *Seminars in Nephrology* 29.5, pp. 497–503.
- Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M. et al. (2011). 'High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations'. *Molecular & Cellular Proteomics* 10.10, pp. M111.010090–M111.010090.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Rabini, R., Moretti, N., Staffolani, R., Salvolini, E., Nanetti, L. et al. (2002). 'Reduced susceptibility to peroxidation of erythrocyte plasma membranes from centenarians'. *Experimental Gerontology* 37.5, pp. 657–663.
- Raczy, C., Petrovski, R., Saunders, C. T., Chorny, I., Kruglyak, S. et al. (2013). 'Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms.' *Bioinformatics (Oxford, England)* 29.16, pp. 2041–3.
- Raffler, J., Friedrich, N., Arnold, M., Kacprowski, T., Rueedi, R. et al. (2015). 'Genome-Wide Association Study with Targeted and Non-targeted NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic Individuality.' *PLoS genetics* 11.9, e1005487.
- Rahilly-Tierney, C. R., Lawler, E. V., Scranton, R. E. and Michael Gaziano, J. (2009). 'Low-density lipoprotein reduction and magnitude of cardiovascular risk reduction'. *Preventive Cardiology* 12.2, pp. 80–87.
- Rahman, I. and MacNee, W. (1999). 'Lung glutathione and oxidative stress: implications in cigarette smoke-induced airway disease.' *The American journal of physiology* 277.6 Pt 1, pp. L1067–88.
- Raitakari, O. T., Juonala, M., Rönkämaa, T., Keltikangas-Järvinen, L., Räsänen, L. et al. (2008). 'Cohort profile: The cardiovascular risk in young Finns study'. *International Journal of Epidemiology* 37.6, pp. 1220–1226.
- Ramezani, A., Massy, Z. A., Meijers, B., Evenepoel, P., Vanholder, R. et al. (2016). 'Role of the Gut Microbiome in Uremia: A Potential Therapeutic Target.' *American journal of kidney diseases : the official journal of the National Kidney Foundation* 67.3, pp. 483–98.
- Ramezani, A. and Raj, D. S. (2014). 'The gut microbiome, kidney disease, and targeted interventions.' *Journal of the American Society of Nephrology : JASN* 25.4, pp. 657–70.
- Rand, W. M. (1971). 'Objective Criteria for the Evaluation of Clustering Methods'. *Journal of the American Statistical Association* 66.336, p. 846.
- Raschenberger, J., Kollerits, B., Ritchie, J., Lane, B., Kalra, P. A. et al. (2015). 'Association of relative telomere length with progression of chronic kidney disease in two cohorts: effect modification by smoking and diabetes'. *Scientific Reports* 5.January, p. 11887.
- Rattan, S. I. S. (2006). 'Theories of biological aging: genes, proteins, and free radicals.' *Free radical research* 40.12, pp. 1230–8.
- Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). 'High-dimensional Ising model selection using l1 -regularized logistic regression'. *EN. The Annals of Statistics* 38.3, pp. 1287–1319.
- Reddy, M. a. and Natarajan, R. (2011). 'Epigenetics in diabetic kidney disease'. *J Am Soc Nephrol* 22.12, pp. 2182–2185.
- Reddy, M. A. and Natarajan, R. (2015). 'Recent developments in epigenetics of acute and chronic kidney diseases'. *Kidney international* 88.2, pp. 250–61.
- Rhee, E. P., Cheng, S., Larson, M. G., Walford, G. A., Lewis, G. D. et al. (2011). 'Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans'. *Journal of Clinical Investigation* 121.4, pp. 1402–1411.

- Rieg, T., Steigle, H., Schnermann, J., Richter, K., Osswald, H. et al. (2005). 'Requirement of intact adenosine A1 receptors for the diuretic and natriuretic action of the methylxanthines theophylline and caffeine.' *The Journal of pharmacology and experimental therapeutics* 313.1, pp. 403–9.
- Rijsdijk, F. V. and Sham, P. C. (2002). 'Analytic approaches to twin data using structural equation models.' *Briefings in bioinformatics* 3.2, pp. 119–33.
- Ríos-Covián, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., Reyes-Gavilán, C. G. de los et al. (2016). 'Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health'. *Frontiers in Microbiology* 7.
- Rockwood, K., Fox, R. A., Stolee, P., Robertson, D. and Beattie, B. L. (1994). 'Frailty in elderly people: An evolving concept'. *Cmaj* 150.4, pp. 489–495.
- Rockwood, K. (2005). 'A global clinical measure of fitness and frailty in elderly people'. *Canadian Medical Association Journal* 173.5, pp. 489–495.
- Rodwell, G. E. J., Sonu, R., Zahn, J. M., Lund, J., Wilhelmy, J. et al. (2004). 'A transcriptional profile of aging in the human kidney.' *PLoS biology* 2.12, e427.
- Rombouts, Y., Ewing, E., Stadt, L. a. van de, Selman, M. H. J., Trouw, L. a. et al. (2015). 'Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis'. *Annals of the Rheumatic Diseases* 74.1, pp. 234–241.
- Round, J. L. and Mazmanian, S. K. (2009). 'The gut microbiota shapes intestinal immune responses during health and disease.' *Nature reviews. Immunology* 9.5, pp. 313–23.
- Royle, L., Campbell, M. P., Radcliffe, C. M., White, D. M., Harvey, D. J. et al. (2008). 'HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software.' *Analytical biochemistry* 376.1, pp. 1–12.
- Russell, W. R., Gratz, S. W., Duncan, S. H., Holtrop, G., Ince, J. et al. (2011). 'High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health'. *American Journal of Clinical Nutrition* 93.5, pp. 1062–1072.
- Saito, H., Ohtomo, T. and Inui, K. (1996). 'Na(+)-dependent uptake of 1,5-anhydro-D-glucitol via the transport systems for D-glucose and D-mannose in the kidney epithelial cell line, LLC-PK1.' *Nihon Jinzo Gakkai shi* 38.10, pp. 435–40.
- Salamone, L. M., Fuerst, T., Visser, M., Kern, M., Lang, T. et al. (2000). 'Measurement of fat mass using DEXA: a validation study in elderly adults.' *Journal of applied physiology (Bethesda, Md. : 1985)* 89.1, pp. 345–52.
- Saphire, E. O. (2001). 'Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design'. *Science* 293.5532, pp. 1155–1159.
- Sarnak, M. J., Levey, A. S., Schoolwerth, A. C., Coresh, J., Culleton, B. et al. (2003). 'Kidney Disease as a Risk Factor for Development of Cardiovascular Disease: A Statement From the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention'. *Circulation* 108.17, pp. 2154–2169.
- Savica, V., Calò, L., Santoro, D., Monardo, P., Granata, A. et al. (2008). 'Salivary phosphate secretion in chronic kidney disease.' *Journal of renal nutrition : the official journal of the Council on Renal Nutrition of the National Kidney Foundation* 18.1, pp. 87–90.
- Schächter, F., Faure-Delanef, L., Guénot, F., Rouger, H., Froguel, P. et al. (1994). 'Genetic associations with human longevity at the APOE and ACE loci.' *Nature genetics* 6.1, pp. 29–32.
- Schäfer, J. and Strimmer, K. (2005). 'A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics.' *Statistical applications in genetics and molecular biology* 4, Article32.
- Schoolwerth, a. C., Smith, B. C. and Culpepper, R. M. (1988). 'Renal gluconeogenesis.' *Mineral and electrolyte metabolism* 14.6, pp. 347–361.
- Schoonjans, K., Staels, B. and Auwerx, J. (1996). 'The peroxisome proliferator activated receptors (PPARs) and their effects on lipid metabolism and adipocyte differentiation.' *Biochimica et biophysica acta* 1302.2, pp. 93–109.
- Sebastiani, P., Solovieff, N., Dewan, A. T., Walsh, K. M., Puca, A. et al. (2012). 'Genetic signatures of exceptional longevity in humans.' *PloS one* 7.1, e29848.
- Seidell, J. C. (2000). 'Obesity, insulin resistance and diabetes—a worldwide epidemic.' *The British journal of nutrition* 83 Suppl 1.S1, S5–8.
- Sekula, P., Goek, O.-N., Quaye, L., Barrios, C., Levey, A. S. et al. (2016). 'A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population.' *Journal of the American Society of Nephrology : JASN* 27.4, pp. 1175–88.
- Sender, R., Fuchs, S. and Milo, R. (2016). 'Revised Estimates for the Number of Human and Bacteria Cells in the Body'. *PLOS Biology* 14.8, e1002533.
- Shah, R. D. and Samworth, R. J. (2013). 'Variable selection with error control: another look at stability selection'. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.1, pp. 55–80.

- Shay, J. W. and Wright, W. E. (2011). 'Role of telomeres and telomerase in cancer'. *Seminars in Cancer Biology* 21.6, pp. 349–353.
- Shen, H., Xu, Y., Lu, J., Ma, C., Zhou, Y. et al. (2016). 'Small dense low-density lipoprotein cholesterol was associated with future cardiovascular events in chronic kidney disease patients'. *BMC Nephrology* 17.1, p. 143.
- Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R. et al. (2014). 'An atlas of genetic influences on human blood metabolites'. *Nature genetics* 46.6, pp. 543–50.
- Shirley, D. G., Walter, S. J. and Noormohamed, F. H. (2002). 'Natriuretic effect of caffeine: assessment of segmental sodium reabsorption in humans'. *Clinical science (London, England : 1979)* 103.5, pp. 461–6.
- Silverstein, D. M. (2009). 'Inflammation in chronic kidney disease: role in the progression of renal and cardiovascular disease'. *Pediatric nephrology (Berlin, Germany)* 24.8, pp. 1445–52.
- Slade, P. G., Caspary, R. G., Nargund, S. and Huang, C. J. (2016). 'Mannose metabolism in recombinant CHO cells and its effect on IgG glycosylation'. *Biotechnology and Bioengineering* 113.7, pp. 1468–1480.
- Smith, D. H. (2004). 'Cost of Medical Care for Chronic Kidney Disease and Comorbidity among Enrollees in a Large HMO Population'. *Journal of the American Society of Nephrology* 15.5, pp. 1300–1306.
- Smith, J. D. (2002). 'Apolipoproteins and aging: emerging mechanisms'. *Ageing research reviews* 1.3, pp. 345–65.
- Smyth, L. J., McKay, G. J., Maxwell, A. P. and McKnight, A. J. (2014). 'DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease'. *Epigenetics* 9.3, pp. 366–376.
- Soininen, P., Kangas, A. J., Würtz, P., Suna, T. and Ala-Korpela, M. (2015). 'Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics'. *Circulation: Cardiovascular Genetics* 8.1, pp. 192–206.
- Speed, D., Hemani, G., Johnson, M. R. and Balding, D. J. (2012). 'Improved Heritability Estimation from Genome-wide SNPs'. *The American Journal of Human Genetics* 91.6, pp. 1011–1021.
- Stein, E. A., Lane, M. and Laskarzewski, P. (1998). 'Comparison of Statins in Hypertriglyceridemia'. *The American Journal of Cardiology* 81.4, 66B–69B.
- Steuer, R. (2006). 'Review: on the analysis and interpretation of correlations in metabolomic data'. *Briefings in bioinformatics* 7.2, pp. 151–8.
- Stevens, L. A. and Levey, A. S. (2009). 'Measured GFR as a confirmatory test for estimated GFR'. *Journal of the American Society of Nephrology : JASN* 20.11, pp. 2305–13.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L. et al. (2005). 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles'. *eng. Proceedings of the National Academy of Sciences of the United States of America* 102.43, pp. 15545–15550.
- Suhre, K., Meisinger, C., Döring, A., Altmaier, E., Belcredi, P. et al. (2010). 'Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting'. *PLoS one* 5.11, e13953.
- Sun, J., Shannon, M., Ando, Y., Schnackenberg, L. K., Khan, N. A. et al. (2012). 'Serum metabolomic profiles from patients with acute kidney injury: A pilot study'. *Journal of Chromatography B* 893–894, pp. 107–113.
- Sun, S., Lourie, R., Cohen, S. B., Ji, Y., Goodrich, J. K. et al. (2016). 'Epithelial Sel1L is required for the maintenance of intestinal homeostasis'. *Molecular biology of the cell* 27.3, pp. 483–90.
- Sun, S., Shi, G., Han, X., Francisco, A. B., Ji, Y. et al. (2014). 'Sel1L is indispensable for mammalian endoplasmic reticulum-associated degradation, endoplasmic reticulum homeostasis, and survival'. *Proceedings of the National Academy of Sciences of the United States of America* 111.5, E582–91.
- Swardfager, W., Lanctôt, K., Rothenburg, L., Wong, A., Cappell, J. et al. (2010). 'A Meta-Analysis of Cytokines in Alzheimer's Disease'. *Biological Psychiatry* 68.10, pp. 930–941.
- Sze, M. A. and Schloss, P. D. (2016). 'Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome'. *mBio* 7.4, e01018–16.
- Szeto, A., Nation, D. a., Mendez, A. J., Dominguez-Bendala, J., Brooks, L. G. et al. (2008). 'Oxytocin attenuates NADPH-dependent superoxide activity and IL-6 secretion in macrophages and vascular cells'. *American journal of physiology. Endocrinology and metabolism* 295.6, E1495–501.
- Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G. et al. (2013). 'Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing'. *Nucleic acids research* 41.Database issue, pp. D1027–33.
- Takahashi, N. (2015). 'Oral Microbiome Metabolism: From "Who Are They?" to "What Are They Doing?".' *Journal of dental research* 94.12, pp. 1628–37.
- Takahashi, S., Yamamoto, T., Tsutsumi, Z., Moriwaki, Y., Yamakita, J. et al. (1997). 'Close correlation between visceral fat accumulation and uric acid metabolism in healthy men'. *Metabolism* 46.10, pp. 1162–1165.
- Tall, A. R. (1998). 'An overview of reverse cholesterol transport'. *European heart journal* 19 Suppl A, A31–5.
- Tang, H.-F., Chen, J.-Q., Xie, Q.-M., Zheng, X.-Y., Zhu, Y.-L. et al. (2006). 'The role of PDE4 in pulmonary inflammation and goblet cell hyperplasia in allergic rats'. *Biochimica et biophysica acta* 1762.5, pp. 525–532.

- Teicher, B. a., Linehan, W. M. and Helman, L. J. (2012). 'Targeting cancer metabolism.' *Clinical cancer research : an official journal of the American Association for Cancer Research* 18.20, pp. 5537–45.
- Telenti, A., Pierce, L. C. T., Biggs, W. H., Iulio, J. di, Wong, E. H. M. et al. (2016). 'Deep sequencing of 10,000 human genomes.' *Proceedings of the National Academy of Sciences of the United States of America* 113.42, pp. 11901–11906.
- Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J. et al. (2013). 'A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.' *Bioinformatics (Oxford, England)* 29.2, pp. 189–96.
- Teucher, B., Skinner, J., Skidmore, P. M. L., Cassidy, A., Fairweather-Tait, S. J. et al. (2007). 'Dietary patterns and heritability of food choice in a UK female twin cohort.' *Twin research and human genetics* 10.5, pp. 734–48.
- Theodoratou, E., Campbell, H., Venthani, N. T., Kolarich, D., Pučić-Baković, M. et al. (2014). 'The role of glycosylation in IBD.' *Nature reviews. Gastroenterology & hepatology* 11.10, pp. 588–600.
- Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S. et al. (2013). 'A community-driven global reconstruction of human metabolism.' *Nature biotechnology* 31.5, pp. 419–25.
- Thomas, C. and Thomas, L. (2009). 'Renal failure—measuring the glomerular filtration rate.' *Deutsches Arzteblatt international* 106.51–52, pp. 849–54.
- Thomas, P., O' Callaghan, N. J. and Fenech, M. (2008). 'Telomere length in white blood cells, buccal cells and brain tissue and its variation with ageing and Alzheimer's disease'. *Mechanisms of Ageing and Development* 129.4, pp. 183–190.
- Thorn, C. F., Aklillu, E., McDonagh, E. M., Klein, T. E. and Altman, R. B. (2012). 'PharmGKB summary: caffeine pathway.' *Pharmacogenetics and genomics* 22.5, pp. 389–95.
- Tibshirani, R. (1994). 'Regression Selection and Shrinkage via the Lasso'. *Journal of the Royal Statistical Society B* 58.1, pp. 267–288.
- Tilg, H. and Moschen, A. R. (2006). 'Adipocytokines: mediators linking adipose tissue, inflammation and immunity.' *Nature reviews. Immunology* 6.10, pp. 772–83.
- Timmeren, M. M. van, Veen, B. S. van der, Stegeman, C. a., Petersen, A. H., Hellmark, T. et al. (2010). 'IgG glycan hydrolysis attenuates ANCA-mediated glomerulonephritis.' *Journal of the American Society of Nephrology : JASN* 21.7, pp. 1103–1114.
- Tizianello, A., De Ferrari, G., Garibotto, G., Gurreri, G. and Robaudo, C. (1980). 'Renal metabolism of amino acids and ammonia in subjects with normal renal function and in patients with chronic renal insufficiency.' *The Journal of clinical investigation* 65.5, pp. 1162–73.
- Tizianello, A., Deferrari, G., Garibotto, G., Robaudo, C., Lutman, M. et al. (1983). 'Branched-chain amino acid metabolism in chronic renal failure.' *Kidney international. Supplement* 16, S17–22.
- Tomana, M., Schrohenloher, R. E., Koopman, W. J., Alarcón, G. S. and Paul, W. A. (1988). 'Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases.' *Arthritis and rheumatism* 31.3, pp. 333–8.
- Tomás, I., Marinho, J. S., Limeres, J., Santos, M. J., Araújo, L. et al. (2008). 'Changes in salivary composition in patients with renal failure.' *Archives of oral biology* 53.6, pp. 528–32.
- Tonelli, M. and Riella, M. (2014). 'Chronic kidney disease and the aging population'. *Indian Journal of Nephrology* 24.2, p. 71.
- Tonelli, M., Sacks, F., Pfeffer, M., Jhangri, G. S. and Curhan, G. (2005). 'Biomarkers of inflammation and progression of chronic kidney disease'. *Kidney International* 68.1, pp. 237–245.
- Toyohara, T., Akiyama, Y., Suzuki, T., Takeuchi, Y., Mishima, E. et al. (2010). 'Metabolomic profiling of uremic solutes in CKD patients'. *Hypertension research : official journal of the Japanese Society of Hypertension* 33.9, pp. 944–952.
- Tsai, P.-C., Van Dongen, J., Tan, Q., Willemsen, G., Christiansen, L. et al. (2015). 'DNA Methylation Changes in the IGF1R Gene in Birth Weight Discordant Adult Monozygotic Twins.' *Twin research and human genetics : the official journal of the International Society for Twin Studies* 18.6, pp. 635–46.
- Tsimihodimos, V., Dounousi, E. and Siamopoulos, K. C. (2008). 'Dyslipidemia in chronic kidney disease: An approach to pathogenesis and treatment'. *American Journal of Nephrology* 28.6, pp. 958–973.
- Tu, T., Gao, Q., Luo, Y., Chen, J., Lu, D. et al. (2013). 'CD146 deletion in the nervous system impairs appetite, locomotor activity and spatial learning in mice.' *PLoS one* 8.9, e74124.
- Tur, I. and Castelo, R. (2012). 'Learning high-dimensional mixed graphical models with missing values'. *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models*. Granada. Chap. Learning h, pp. 323–330.
- Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A. et al. (2009). 'A core gut microbiome in obese and lean twins'. *Nature* 457.7228, pp. 480–484.
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. et al. (2007). 'The human microbiome project.' *Nature* 449.7164, pp. 804–10.

- United States Renal Data System (2016). *USRDS 2016 Annual Data Report*. Tech. rep. Ann Arbor, MI: United States Renal Data System.
- Valcárcel, B., Würtz, P., Seich al Basatena, N.-K., Tukiainen, T., Kangas, A. J. et al. (2011). 'A Differential Network Approach to Exploring Differences between Biological States: An Application to Prediabetes'. *PLoS ONE* 6.9. Ed. by Csermely, P., e24702.
- Valdes, A. M., Andrew, T., Gardner, J. P., Kimura, M., Oelsner, E. et al. (2005). 'Obesity, cigarette smoking, and telomere length in women'. *Lancet* 366.9486, pp. 662–664.
- Valdes, A. M., Glass, D. and Spector, T. D. (2013). 'Omics technologies and the study of human ageing'. *Nature reviews. Genetics* 14.9, pp. 601–7.
- Väremo, L., Nielsen, J. and Nookaew, I. (2013). 'Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods'. *Nucleic Acids Research* 41.8, pp. 4378–4391.
- Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P. et al. (2009). *Essentials of Glycobiology*. Ed. by Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P. et al. 2nd. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, p. 800.
- Vaziri, N. D., Wong, J., Pahl, M., Piceno, Y. M., Yuan, J. et al. (2013). 'Chronic kidney disease alters intestinal microbial flora'. *Kidney international* 83.2, pp. 308–15.
- Vehtari, A., Mäkinen, V.-P., Soininen, P., Ingman, P., Mäkelä, S. M. et al. (2007). 'A novel Bayesian approach to quantify clinical variables and to determine their spectroscopic counterparts in 1H NMR metabonomic data'. *BMC bioinformatics* 8 Suppl 2.Suppl 2, S8.
- Vera, E., Bernardes de Jesus, B., Foronda, M., Flores, J. M. and Blasco, M. a. (2012). 'The rate of increase of short telomeres predicts longevity in mammals'. *Cell reports* 2.4, pp. 732–7.
- Verrey, F., Singer, D., Ramadan, T., Vuille-dit-Bille, R. N., Mariotta, L. et al. (2009). 'Kidney amino acid transport'. *Pflugers Archiv : European journal of physiology* 458.1, pp. 53–60.
- Vijg, J. and Suh, Y. (2005). 'Genetics of longevity and aging'. *Annual review of medicine* 56, pp. 193–212.
- Viña, J., Borrás, C. and Miquel, J. (2007). 'Theories of ageing'. *IUBMB Life* 59.4, pp. 249–254.
- Vogl, F. C., Mehrl, S., Heizinger, L., Schlecht, I., Zacharias, H. U. et al. (2016). 'Evaluation of dilution and normalization strategies to correct for urinary output in HPLC-HRTOFMS metabolomics'. *Analytical and bioanalytical chemistry* 408.29, pp. 8483–8493.
- Völzke, H., Alte, D., Schmidt, C. O., Radke, D., Lohrer, R. et al. (2011). 'Cohort profile: the study of health in Pomerania'. *International journal of epidemiology* 40.2, pp. 294–307.
- Vučković, F., Krištic, J., Gudelj, I., Teruel, M., Keser, T. et al. (2015). 'Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome'. *Arthritis and Rheumatology* 67.11, pp. 2978–2989.
- Wainwright, M. J., Ravikumar, P. and Lafferty, J. D. (2006). 'High-Dimensional Graphical Model Selection Using l1-Regularized Logistic Regression'. *Neural Information Processing Systems*. Vancouver.
- Walker, B. R. (2007). 'Glucocorticoids and cardiovascular disease'. *European Journal of Endocrinology* 157.5, pp. 545–559.
- Wang, J., Zhang, S., Wang, Y., Chen, L. and Zhang, X.-S. (2009). 'Disease-aging network reveals significant roles of aging genes in connecting genetic diseases'. *PLoS computational biology* 5.9, e1000521.
- Wang, K., Li, W.-D., Zhang, C. K., Wang, Z., Glessner, J. T. et al. (2011a). 'A genome-wide association study on obesity and obesity-related traits'. *PloS one* 6.4. Ed. by Zhao, Z., e18939.
- Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P. et al. (2011b). 'Metabolite profiles and the risk of developing diabetes'. *Nature medicine* 17.4, pp. 448–53.
- Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F. et al. (2014). 'Aging of blood can be tracked by DNA methylation changes at just three CpG sites'. *Genome biology* 15.2, R24.
- Weimann, A., Sabroe, M. and Poulsen, H. E. (2005). 'Measurement of caffeine and five of the major metabolites in urine by high-performance liquid chromatography/tandem mass spectrometry'. *Journal of mass spectrometry : JMS* 40.3, pp. 307–16.
- Weinert, B. T. and Timiras, P. S. (2003). 'Invited review: Theories of aging'. *Journal of applied physiology (Bethesda, Md. : 1985)* 95.4, pp. 1706–16.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P. et al. (2014). 'The NHGRI GWAS Catalog, a curated resource of SNP-trait associations'. *Nucleic acids research* 42.Database issue, pp. D1001–6.
- West, J., Beck, S., Wang, X. and Teschendorff, A. E. (2013). 'An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways'. *Scientific reports* 3, p. 1630.
- Westcott, S. L. and Schloss, P. D. (2015). 'De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units'. *PeerJ* 3, e1487.
- Whitfield, J. B. (2001). 'Gamma glutamyl transferase'. eng. *Critical reviews in clinical laboratory sciences* 38.4, pp. 263–355.

- Wichmann, H.-E., Gieger, C. and Illig, T. (2005). 'KORA-gen - Resource for Population Genetics, Controls and a Broad Spectrum of Disease Phenotypes'. *Das Gesundheitswesen* 67.S 01, pp. 26–30.
- Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M. et al. (2014). 'Mass spectrometry based draft of the human proteome.' *Nature* 509.7502, pp. 582–7.
- Willcox, D. C., Willcox, B. J. and Poon, L. W. (2010). 'Centenarian Studies: Important Contributors to Our Understanding of the Aging Process and Longevity'. *Current Gerontology and Geriatrics Research* 2010, pp. 1–6.
- Wing, M. R., Devaney, J. M., Joffe, M. M., Xie, D., Feldman, H. I. et al. (2014). 'DNA methylation profile associated with rapid decline in kidney function: Findings from the CRIC Study'. *Nephrology Dialysis Transplantation* 29.4, pp. 864–872.
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C. et al. (2013). 'HMDB 3.0—The Human Metabolome Database in 2013.' *Nucleic acids research* 41.Database issue, pp. D801–7.
- World Health Organization (2016). *Global Report on Diabetes*. Tech. rep. World Health Organisation, p. 88.
- Wu, G., Fang, Y.-Z., Yang, S., Lupton, J. R. and Turner, N. D. (2004). 'Glutathione metabolism and its implications for health.' *The Journal of nutrition* 134.3, pp. 489–492.
- Wurtz, P., Havulinna, A. S., Soininen, P., Tynkkynen, T., Prieto-Merino, D. et al. (2015). 'Metabolite Profiling and Cardiovascular Event Risk: A Prospective Study of 3 Population-Based Cohorts'. *Circulation* 131.9, pp. 774–785.
- Wurtz, P., Kangas, A., Soininen, P., Lawlor, D., Davey Smith, G. et al. (2017). 'Quantitative serum NMR metabolomics in large-scale epidemiology a primer on -omic technology'. *American Journal of Epidemiology*.
- Würtz, P., Wang, Q., Soininen, P., Kangas, A. J., Fatemifar, G. et al. (2016). 'Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase.' *Journal of the American College of Cardiology* 67.10, pp. 1200–10.
- Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S.-M. et al. (2002). 'DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions.' *Nucleic acids research* 30.1, pp. 303–5.
- Xia, K., Xue, H., Dong, D., Zhu, S., Wang, J. et al. (2006). 'Identification of the proliferation / differentiation switch in the cellular network of multicellular organisms.' *PLoS computational biology* 2.11, e145.
- Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z. et al. (2016). 'Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome'. *Cell Systems* 3.6, 572–584.e3.
- Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S. et al. (2007). 'A modular network model of aging.' *Molecular systems biology* 3.147, p. 147.
- Yamanouchi, T. and Akanuma, Y. (1994). 'Serum 1,5-anhydroglucitol (1,5 AG): new clinical marker for glycemic control.' *Diabetes research and clinical practice* 24 Suppl, S261–8.
- Yang, J., Lee, S. H., Goddard, M. E. and Visscher, P. M. (2011). 'GCTA: a tool for genome-wide complex trait analysis.' *American journal of human genetics* 88.1, pp. 76–82.
- Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H. and Qu, L.-H. (2013). 'ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data.' *Nucleic acids research* 41.Database issue, pp. D177–87.
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G. et al. (2012). 'Human gut microbiome viewed across age and geography'. *Nature* 486.7402, pp. 222–227.
- Yazdani, A., Yazdani, A., Samiei, A. and Boerwinkle, E. (2016). 'Generating a robust statistical causal structure over 13 cardiovascular disease risk factors using genomics data'. *Journal of Biomedical Informatics* 60, pp. 114–119.
- Yokozawa, T., Mo, Z. L. and Oura, H. (1989). 'Comparison of toxic effects of methylguanidine, guanidinosuccinic acid and creatinine in rats with adenine-induced chronic renal failure.' *Nephron* 51.3, pp. 388–92.
- Young, G. A. and Parsons, F. M. (1973). 'Impairment of phenylalanine hydroxylation in chronic renal insufficiency.' *Clinical science* 45.1, pp. 89–97.
- Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. and Gerstein, M. (2007). 'The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics.' *PLoS computational biology* 3.4, e59.
- Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T. et al. (2012). 'Human serum metabolic profiles are age dependent.' *Aging cell* 11.6, pp. 960–7.
- Yuan, M. and Lin, Y. (2007). 'Model selection and estimation in the Gaussian graphical model'. *Biometrika* 94.1, pp. 19–35.
- Yuan, Y., Li, C.-T. and Windram, O. (2011). 'Directed partial correlation: inferring large-scale gene regulatory network through induced topology disruptions.' *PloS one* 6.4, e16835.
- Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A. et al. (2013). 'Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation'. *PLoS ONE* 8.5, e63812.

- Zhai, G., Aviv, A., Hunter, D. J., Hart, D. J., Gardner, J. P. et al. (2006). 'Reduction of leucocyte telomere length in radiographic hand osteoarthritis: a population-based study.' *Annals of the rheumatic diseases* 65.11, pp. 1444–1448.
- Zhang, B. and Horvath, S. (2005a). 'A general framework for weighted gene co-expression network analysis.' *Statistical applications in genetics and molecular biology* 4.1, Article17.
- Zhang, H., Wu, C., Chen, Q., Chen, X., Xu, Z. et al. (2013). 'Treatment of Obesity and Diabetes Using Oxytocin or Analogs in Patients and Mouse Models'. *PLoS ONE* 8.5. Ed. by Wagner, B., e61477.
- Zhang, H., Forman, H. J. and Choi, J. (2005b). 'Gamma-Glutamyl Transpeptidase in Glutathione Biosynthesis'. *Methods in Enzymology* 401.05, pp. 468–483.
- Zhang, Y., Zhang, S. and Wang, G. (2015). 'Metabolomic biomarkers in diabetic kidney diseases—A systematic review.' *Journal of diabetes and its complications* 29.8, pp. 1345–51.
- Zhao, J., Zhu, Y., Uppal, K., Tran, V. T., Yu, T. et al. (2014). 'Metabolic profiles of biological aging in american indians: The strong heart family study'. *Aging* 6.3, pp. 176–186.
- Zhou, H., Tan, K. C. B., Shiu, S. W. M. and Wong, Y. (2008). 'Cellular cholesterol efflux to serum is impaired in diabetic nephropathy.' *Diabetes/metabolism research and reviews* 24.8, pp. 617–23.
- Zhou, X. and Stephens, M. (2012). 'Genome-wide efficient mixed-model analysis for association studies'. *Nature genetics* 44.7, pp. 821–4.
- Zhu, B., Wang, X. and Li, L. (2010). 'Human gut microbiome: the second genome of human body.' *Protein & cell* 1.8, pp. 718–25.
- Zierer, J., Kastenmüller, G., Suhre, K., Gieger, C., Codd, V. et al. (2016a). 'Metabolomics profiling reveals novel markers for leukocyte telomere length.' *Aging* 8.1, pp. 77–94.
- Zierer, J., Menni, C., Kastenmüller, G. and Spector, T. D. (2015). 'Integration of 'omics' data in aging research: from biomarkers to systems biology'. *Aging Cell* 14.6, pp. 933–944.
- Zierer, J., Pallister, T., Tsai, P.-C., Krumsiek, J., Bell, J. T. et al. (2016b). 'Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model'. *Scientific Reports* 6.1, p. 37646.
- Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y. et al. (2013). 'Charting a dynamic DNA methylation landscape of the human genome.' *Nature* 500.7463, pp. 477–81.
- Ziller, M. J., Hansen, K. D., Meissner, A. and Aryee, M. J. (2014). 'Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing'. *Nature Methods* 12.3, pp. 230–232.
- Zomer, a. W., van Der Burg, B., Jansen, G. a., Wanders, R. J., Poll-The, B. T. et al. (2000). 'Pristanic acid and phytanic acid: naturally occurring ligands for the nuclear receptor peroxisome proliferator-activated receptor alpha.' *Journal of lipid research* 41.11, pp. 1801–7.

Appendix

APPENDIX A

Metabolic markers of telomere length

Table A.1 Associations of circulating metabolites with telomere length in TwinsUK cohort.

Associations of 280 circulating metabolites with leukocyte telomere length (LTL) were analysed. Metabolites are sorted by the strength of their association with LTL. Metabolites passing Bonferroni correction are indicated by **, nominally significant associations by *.

Metabolite	Pathway	β [95 %CI]	p-value
1-stearoylglycerophosphoinositol	Lipid	-0.09 [-0.13;-0.05]	$1.4 \times 10^{-6} **$
gamma-glutamyltyrosine	Peptide	-0.09 [-0.12;-0.05]	$3.4 \times 10^{-6} **$
gamma-glutamylphenylalanine	Peptide	-0.08 [-0.12;-0.04]	$2.7 \times 10^{-5} **$
1-palmitoylglycerophosphoinositol	Lipid	-0.08 [-0.13;-0.04]	$7.4 \times 10^{-5} **$
4-vinylphenol sulfate	Xenobiotics	-0.08 [-0.12;-0.04]	$7.4 \times 10^{-5} **$
1-arachidonoylglycerophosphoethanolamine	Lipid	-0.07 [-0.11;-0.03]	$2.4 \times 10^{-4} *$
1-arachidonoylglycerophosphoinositol	Lipid	-0.07 [-0.11;-0.03]	$2.8 \times 10^{-4} *$
1-oleoylglycerophosphoethanolamine	Lipid	-0.07 [-0.10;-0.03]	$7.6 \times 10^{-4} *$
uridine	Nucleotide	0.06 [0.03: 0.10]	$1.1 \times 10^{-3} *$
caprylate (8:0)	Lipid	-0.06 [-0.10;-0.02]	$1.4 \times 10^{-3} *$
erythritol	Xenobiotics	-0.06 [-0.09;-0.02]	$1.5 \times 10^{-3} *$
arabinose	Carbohydrate	-0.07 [-0.11;-0.02]	$1.8 \times 10^{-3} *$
ornithine	Amino acid	-0.06 [-0.10;-0.02]	$2.1 \times 10^{-3} *$
heptanoate (7:0)	Lipid	-0.06 [-0.09;-0.02]	3.3×10^{-3}
pyroglutamine	Amino acid	-0.05 [-0.09;-0.02]	3.5×10^{-3}
tyrosine	Amino acid	-0.05 [-0.09;-0.02]	5.3×10^{-3}
laurylcarnitine	Lipid	-0.07 [-0.11;-0.02]	6.7×10^{-3}
proline	Amino acid	-0.05 [-0.09;-0.01]	7.6×10^{-3}
malate	Energy	-0.05 [-0.09;-0.01]	7.7×10^{-3}
caproate (6:0)	Lipid	-0.05 [-0.09;-0.01]	7.7×10^{-3}
kynurenine	Amino acid	-0.05 [-0.08;-0.01]	1.0×10^{-2}
pseudouridine	Nucleotide	-0.05 [-0.08;-0.01]	1.3×10^{-2}
1-palmitoylglycerophosphoethanolamine	Lipid	-0.05 [-0.08;-0.01]	1.3×10^{-2}
arachidonate (20:4n6)	Lipid	-0.05 [-0.09;-0.01]	1.3×10^{-2}
3-dehydrocarnitine	Lipid	-0.05 [-0.08;-0.01]	1.4×10^{-2}
C-glycosyltryptophan	Amino acid	-0.04 [-0.08;-0.01]	1.4×10^{-2}
1,3,7-trimethylurate	Xenobiotics	-0.06 [-0.11;-0.01]	1.4×10^{-2}
4-methyl-2-oxopentanoate	Amino acid	0.05 [0.01: 0.08]	1.6×10^{-2}
bilirubin (E,E)	Cofactors and vitamins	0.05 [0.01: 0.08]	1.6×10^{-2}
p-acetamidophenylglucuronide	Xenobiotics	-0.09 [-0.17;-0.02]	1.6×10^{-2}
glycerol	Lipid	-0.05 [-0.08;-0.01]	1.7×10^{-2}
1-eicosatrienoylglycerophosphocholine	Lipid	-0.05 [-0.08;-0.01]	1.7×10^{-2}
lactate	Carbohydrate	-0.05 [-0.08;-0.01]	1.8×10^{-2}
thromboxane B2	Lipid	-0.04 [-0.08;-0.01]	2.3×10^{-2}
1-palmitoylglycerol (1-monopalmitin)	Lipid	-0.04 [-0.08;-0.01]	2.3×10^{-2}
butyrylcarnitine	Lipid	-0.04 [-0.08;-0.01]	2.4×10^{-2}
2-hydroxyglutarate	Lipid	-0.05 [-0.10;-0.01]	2.4×10^{-2}
homocitrulline	Amino acid	-0.06 [-0.12;-0.01]	2.7×10^{-2}
phenylalanine	Amino acid	-0.04 [-0.08;-0.00]	2.7×10^{-2}
alpha-hydroxyisovalerate	Amino acid	-0.04 [-0.08;-0.00]	3.0×10^{-2}
CMPF	Lipid	0.04 [0.00: 0.07]	3.2×10^{-2}
aspartate	Amino acid	-0.04 [-0.08;-0.00]	3.2×10^{-2}
theophylline	Xenobiotics	-0.04 [-0.08;-0.00]	3.4×10^{-2}
biliverdin	Cofactors and vitamins	0.04 [0.00: 0.08]	3.5×10^{-2}
cholesterol	Lipid	-0.04 [-0.07;-0.00]	3.6×10^{-2}
oleoylcarnitine	Lipid	0.04 [0.00: 0.08]	3.8×10^{-2}
pelargonate (9:0)	Lipid	-0.04 [-0.07;-0.00]	4.0×10^{-2}
paraxanthine	Xenobiotics	-0.04 [-0.08;-0.00]	4.1×10^{-2}

continued on next page ...

3-(cystein-S-yl)acetaminophen	Xenobiotics	-0.09[-0.18:-0.00]	4.2×10^{-2}
urate	Nucleotide	-0.04[-0.07:-0.00]	4.2×10^{-2}
carnitine	Lipid	-0.04[-0.08:-0.00]	4.3×10^{-2}
1-linoleoylglycerophosphoethanolamine	Lipid	-0.04[-0.08:-0.00]	4.4×10^{-2}
1-stearoylglycerophosphoethanolamine	Lipid	-0.04[-0.08:-0.00]	4.5×10^{-2}
adrenate (22:4n6)	Lipid	-0.04[-0.08:-0.00]	4.5×10^{-2}
gamma-glutamylvaline	Peptide	-0.04[-0.07:-0.00]	4.8×10^{-2}
threonate	Cofactors and vitamins	0.04[0.00:0.08]	5.0×10^{-2}
tetradecanedioate	Lipid	-0.04[-0.08:0.00]	5.1×10^{-2}
hydoxycholate	Lipid	-0.04[-0.08:0.00]	5.1×10^{-2}
2-hydroxystearate	Lipid	-0.04[-0.07:0.00]	5.2×10^{-2}
ibuprofen	Xenobiotics	-0.07[-0.13:0.00]	5.2×10^{-2}
2-methoxyacetaminophen sulfate	Xenobiotics	-0.08[-0.17:0.00]	5.2×10^{-2}
2-hydroxypalmitate	Lipid	-0.04[-0.07:0.00]	5.2×10^{-2}
hexadecanedioate	Lipid	-0.04[-0.08:0.00]	5.4×10^{-2}
gamma-glutamylthreonine	Peptide	-0.06[-0.11:0.00]	5.5×10^{-2}
alanine	Amino acid	-0.04[-0.07:0.00]	5.7×10^{-2}
dihomo-linolenate (20:3n3 or n6)	Lipid	-0.04[-0.07:0.00]	5.9×10^{-2}
undecanoate (11:0)	Lipid	-0.04[-0.07:0.00]	6.0×10^{-2}
glutamate	Amino acid	-0.03[-0.07:0.00]	6.1×10^{-2}
glycerate	Carbohydrate	0.03[-0.00:0.07]	7.3×10^{-2}
ADpSGEGDFXAEGGVR	Peptide	0.06[-0.01:0.12]	7.4×10^{-2}
naproxen	Xenobiotics	-0.33[-0.68:0.02]	8.2×10^{-2}
saccharin	Xenobiotics	-0.06[-0.12:0.01]	8.3×10^{-2}
aspartylphenylalanine	Peptide	-0.05[-0.12:0.01]	8.3×10^{-2}
beta-hydroxyisovalerate	Amino acid	-0.03[-0.07:0.00]	8.5×10^{-2}
propionylcarnitine	Lipid	-0.03[-0.07:0.00]	8.6×10^{-2}
7-Hoca	Lipid	-0.03[-0.07:0.00]	8.7×10^{-2}
cyclo(leu-pro)	Peptide	-0.04[-0.09:0.01]	8.7×10^{-2}
estrone 3-sulfate	Lipid	-0.09[-0.18:0.01]	8.9×10^{-2}
benzoate	Xenobiotics	-0.03[-0.07:0.01]	9.3×10^{-2}
1-methylxanthine	Xenobiotics	-0.04[-0.08:0.01]	9.3×10^{-2}
N-acetylalanine	Amino acid	-0.03[-0.07:0.01]	9.6×10^{-2}
pyridoxate	Cofactors and vitamins	0.03[-0.01:0.07]	9.7×10^{-2}
4-acetamidobutanoate	Amino acid	-0.03[-0.07:0.01]	9.8×10^{-2}
gamma-glutamylleucine	Peptide	-0.03[-0.07:0.01]	1.0×10^{-1}
4-androsten-3beta,17beta-diol disulfate 1	Lipid	-0.03[-0.07:0.01]	1.1×10^{-1}
1-arachidonoylglycerophosphocholine	Lipid	-0.03[-0.07:0.01]	1.1×10^{-1}
1-palmitoylplasmenylethanolamine	Lipid	-0.03[-0.08:0.01]	1.1×10^{-1}
2-hydroxyisobutyrate	Amino acid	0.03[-0.01:0.07]	1.1×10^{-1}
quinate	Xenobiotics	-0.03[-0.08:0.01]	1.1×10^{-1}
caffeine	Xenobiotics	-0.03[-0.07:0.01]	1.1×10^{-1}
3-methyl-2-oxovalerate	Amino acid	0.03[-0.01:0.07]	1.2×10^{-1}
4-acetamidophenol	Xenobiotics	-0.08[-0.18:0.02]	1.2×10^{-1}
N2,N2-dimethylguanosine	Nucleotide	-0.03[-0.07:0.01]	1.3×10^{-1}
cholate	Lipid	0.03[-0.01:0.08]	1.4×10^{-1}
succinylcarnitine	Energy	-0.03[-0.07:0.01]	1.4×10^{-1}
stachydrine	Xenobiotics	0.03[-0.01:0.07]	1.4×10^{-1}
citrate	Energy	0.03[-0.01:0.06]	1.4×10^{-1}
1-palmitoleoylglycerophosphocholine	Lipid	-0.03[-0.07:0.01]	1.4×10^{-1}
gamma-glutamylisoleucine	Peptide	-0.03[-0.07:0.01]	1.5×10^{-1}
phenyllactate	Amino acid	-0.03[-0.07:0.01]	1.5×10^{-1}
hydroxyisovaleroyl carnitine	Amino acid	-0.03[-0.08:0.01]	1.5×10^{-1}
isoleucine	Amino acid	-0.03[-0.06:0.01]	1.6×10^{-1}
palmitate (16:0)	Lipid	-0.03[-0.07:0.01]	1.6×10^{-1}
1,7-dimethylurate	Xenobiotics	-0.03[-0.07:0.01]	1.6×10^{-1}
creatinine	Amino acid	-0.03[-0.06:0.01]	1.6×10^{-1}
3-(3-hydroxyphenyl)propionate	Amino acid	0.07[-0.03:0.16]	1.6×10^{-1}
4-hydroxyhippurate	Xenobiotics	-0.03[-0.08:0.01]	1.6×10^{-1}
N-acetylorithine	Amino acid	-0.03[-0.07:0.01]	1.8×10^{-1}
N-(2-furoyl)glycine	Xenobiotics	-0.10[-0.25:0.05]	1.9×10^{-1}
palmitoylcarnitine	Lipid	0.02[-0.01:0.06]	2.0×10^{-1}
1-heptadecanoylglycerophosphocholine	Lipid	0.03[-0.01:0.06]	2.1×10^{-1}
histidine	Amino acid	0.02[-0.01:0.06]	2.1×10^{-1}
4-ethylphenylsulfate	Xenobiotics	-0.03[-0.07:0.02]	2.1×10^{-1}
3-methyl-2-oxobutyrate	Amino acid	0.02[-0.01:0.06]	2.1×10^{-1}
phenol sulfate	Amino acid	-0.02[-0.06:0.01]	2.1×10^{-1}
1,5-anhydroglucitol (1,5-AG)	Carbohydrate	-0.02[-0.06:0.01]	2.1×10^{-1}
alpha-tocopherol	Cofactors and vitamins	0.02[-0.01:0.06]	2.2×10^{-1}
2-hydroxyacetaminophen sulfate	Xenobiotics	-0.04[-0.12:0.03]	2.2×10^{-1}
mannose	Carbohydrate	-0.02[-0.06:0.01]	2.3×10^{-1}
glycochenodeoxycholate	Lipid	-0.02[-0.06:0.02]	2.4×10^{-1}
piperine	Xenobiotics	0.02[-0.02:0.06]	2.4×10^{-1}
deoxycholate	Lipid	-0.03[-0.08:0.02]	2.4×10^{-1}
nonadecanoate (19:0)	Lipid	0.02[-0.01:0.06]	2.4×10^{-1}
hippurate	Xenobiotics	0.02[-0.02:0.06]	2.4×10^{-1}
3-indoxyl sulfate	Amino acid	-0.02[-0.06:0.02]	2.5×10^{-1}
bilirubin (Z,Z)	Cofactors and vitamins	0.02[-0.02:0.06]	2.6×10^{-1}
myristate (14:0)	Lipid	-0.02[-0.06:0.02]	2.6×10^{-1}

continued on next page ...

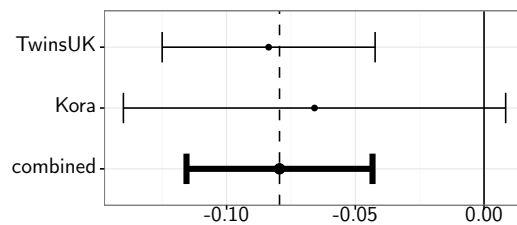
hexanoylcarnitine	Lipid	-0.02[-0.06: 0.02]	2.6×10^{-1}
2-tetradecenoyl carnitine	Lipid	-0.02[-0.06: 0.02]	2.6×10^{-1}
myo-inositol	Lipid	-0.02[-0.06: 0.02]	2.7×10^{-1}
palmitoyl sphingomyelin	Lipid	-0.02[-0.06: 0.02]	2.7×10^{-1}
palmitoleate (16:1n7)	Lipid	-0.02[-0.06: 0.02]	2.7×10^{-1}
ursodeoxycholate	Lipid	-0.02[-0.07: 0.02]	2.8×10^{-1}
homostachydrine	Xenobiotics	-0.04[-0.11: 0.03]	2.8×10^{-1}
glycoursodeoxycholate	Lipid	-0.02[-0.07: 0.02]	2.8×10^{-1}
lathosterol	Lipid	-0.02[-0.06: 0.02]	2.8×10^{-1}
laurate (12:0)	Lipid	-0.02[-0.06: 0.02]	2.9×10^{-1}
3-phenylpropionate (hydrocinnamate)	Amino acid	0.02[-0.02: 0.06]	2.9×10^{-1}
glutaroyl carnitine	Amino acid	-0.02[-0.06: 0.02]	2.9×10^{-1}
DHA (22:6n3)	Lipid	0.02[-0.02: 0.06]	2.9×10^{-1}
bradykinin, des-arg(9)	Peptide	0.03[-0.02: 0.08]	3.0×10^{-1}
7-methylxanthine	Xenobiotics	0.02[-0.02: 0.07]	3.0×10^{-1}
erythrose	Carbohydrate	0.02[-0.02: 0.06]	3.1×10^{-1}
epiandrosterone sulfate	Lipid	-0.02[-0.05: 0.02]	3.1×10^{-1}
1-stearoylglycerol (1-monostearin)	Lipid	-0.02[-0.06: 0.02]	3.1×10^{-1}
10-heptadecenoate (17:1n7)	Lipid	-0.02[-0.06: 0.02]	3.2×10^{-1}
glucose	Carbohydrate	-0.02[-0.06: 0.02]	3.4×10^{-1}
myristoleate (14:1n5)	Lipid	-0.02[-0.06: 0.02]	3.4×10^{-1}
10-undecenoate (11:1n1)	Lipid	0.02[-0.02: 0.06]	3.4×10^{-1}
threonine	Amino acid	-0.02[-0.06: 0.02]	3.4×10^{-1}
taurochenodeoxycholate	Lipid	-0.02[-0.06: 0.02]	3.5×10^{-1}
salicyluric glucuronide	Xenobiotics	-0.06[-0.19: 0.07]	3.5×10^{-1}
5alpha-androstan-3beta,17beta-diol disulfate	Lipid	-0.02[-0.06: 0.02]	3.5×10^{-1}
1-methylurate	Xenobiotics	-0.02[-0.06: 0.02]	3.6×10^{-1}
dodecanedioate	Lipid	-0.02[-0.06: 0.02]	3.6×10^{-1}
GPC	Lipid	-0.02[-0.06: 0.02]	3.6×10^{-1}
tryptophan	Amino acid	0.02[-0.02: 0.06]	3.6×10^{-1}
1-oleoylglycerol (1-monoolein)	Lipid	-0.02[-0.07: 0.02]	3.6×10^{-1}
catechol sulfate	Xenobiotics	-0.02[-0.06: 0.02]	3.6×10^{-1}
mannitol	Carbohydrate	-0.02[-0.06: 0.02]	3.7×10^{-1}
pantothenate	Cofactors and vitamins	0.02[-0.02: 0.06]	3.7×10^{-1}
gamma-tocopherol	Cofactors and vitamins	-0.02[-0.06: 0.02]	3.8×10^{-1}
asparagine	Amino acid	0.02[-0.02: 0.05]	3.8×10^{-1}
1-stearoylglycerophosphocholine	Lipid	0.02[-0.02: 0.05]	3.8×10^{-1}
lysine	Amino acid	-0.02[-0.05: 0.02]	3.8×10^{-1}
stearoylcarnitine	Lipid	0.02[-0.02: 0.05]	3.9×10^{-1}
2-stearoylglycerophosphocholine	Lipid	0.02[-0.02: 0.05]	3.9×10^{-1}
2-methylbutyrylcarnitine	Amino acid	-0.02[-0.05: 0.02]	4.0×10^{-1}
1,6-anhydroglucose	Carbohydrate	-0.02[-0.07: 0.03]	4.1×10^{-1}
glycocholate	Lipid	0.02[-0.02: 0.06]	4.1×10^{-1}
leucine	Amino acid	-0.02[-0.05: 0.02]	4.1×10^{-1}
15-methylpalmitate (isobar with 2-methylpalmitate)	Lipid	-0.02[-0.06: 0.03]	4.1×10^{-1}
acetylphosphate	Energy	-0.01[-0.05: 0.02]	4.3×10^{-1}
5alpha-pregnan-3beta,20alpha-diol disulfate	Lipid	0.02[-0.03: 0.06]	4.4×10^{-1}
cysteine	Amino acid	-0.01[-0.05: 0.02]	4.4×10^{-1}
phenylacetate	Amino acid	0.02[-0.03: 0.06]	4.5×10^{-1}
n-Butyl Oleate	Lipid	0.02[-0.03: 0.07]	4.6×10^{-1}
cis-4-decenoyl carnitine	Lipid	-0.02[-0.06: 0.03]	4.6×10^{-1}
creatine	Amino acid	0.01[-0.02: 0.05]	4.7×10^{-1}
3-(4-hydroxyphenyl)lactate	Amino acid	-0.01[-0.05: 0.02]	4.7×10^{-1}
3-methylhistidine	Amino acid	-0.01[-0.06: 0.03]	4.8×10^{-1}
7-methylguanine	Nucleotide	-0.02[-0.06: 0.03]	4.9×10^{-1}
citrulline	Amino acid	-0.01[-0.05: 0.02]	4.9×10^{-1}
DHEA-S	Lipid	-0.01[-0.04: 0.02]	5.0×10^{-1}
pentadecanoate (15:0)	Lipid	-0.01[-0.05: 0.03]	5.1×10^{-1}
5-dodecenoate (12:1n7)	Lipid	-0.01[-0.05: 0.03]	5.1×10^{-1}
4-androsten-3beta,17beta-diol disulfate 2	Lipid	-0.01[-0.05: 0.03]	5.2×10^{-1}
arginine	Amino acid	0.01[-0.03: 0.05]	5.2×10^{-1}
linoleate (18:2n6)	Lipid	-0.01[-0.05: 0.03]	5.3×10^{-1}
octanoylcarnitine	Lipid	-0.01[-0.05: 0.03]	5.3×10^{-1}
atenolol	Xenobiotics	-0.06[-0.25: 0.13]	5.4×10^{-1}
phenylalanylphenylalanine	Peptide	0.02[-0.04: 0.08]	5.4×10^{-1}
tryptophan betaine	Amino acid	0.01[-0.03: 0.05]	5.4×10^{-1}
androsterone sulfate	Lipid	-0.01[-0.05: 0.02]	5.4×10^{-1}
N-acetylthreonine	Amino acid	-0.01[-0.05: 0.03]	5.5×10^{-1}
2-hydroxyhippurate (salicylurate)	Xenobiotics	0.02[-0.04: 0.08]	5.7×10^{-1}
glycerol 2-phosphate	Xenobiotics	0.01[-0.03: 0.05]	5.9×10^{-1}
dihomo-linoleate (20:2n6)	Lipid	-0.01[-0.05: 0.03]	5.9×10^{-1}
valine	Amino acid	-0.01[-0.05: 0.03]	5.9×10^{-1}
carbamazepine	Xenobiotics	0.12[-0.31: 0.55]	6.0×10^{-1}
3-hydroxybutyrate (BHBA)	Lipid	0.01[-0.03: 0.05]	6.0×10^{-1}
N1-methyladenosine	Nucleotide	-0.01[-0.05: 0.03]	6.2×10^{-1}
betaine	Amino acid	0.01[-0.03: 0.05]	6.2×10^{-1}
decanoylcarnitine	Lipid	0.01[-0.03: 0.05]	6.2×10^{-1}
N-acetylglycine	Amino acid	0.01[-0.03: 0.05]	6.3×10^{-1}
isovalerate	Lipid	-0.01[-0.05: 0.03]	6.3×10^{-1}

continued on next page ...

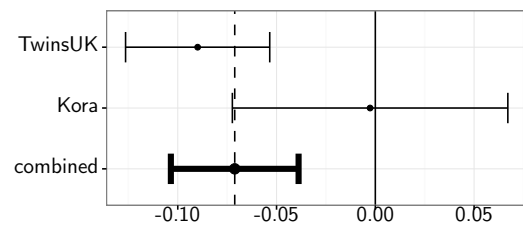
allantoin	Nucleotide	-0.01[-0.05: 0.03]	6.4×10^{-1}
chiro-inositol	Lipid	0.02[-0.05: 0.08]	6.4×10^{-1}
salicylate	Xenobiotics	-0.01[-0.08: 0.05]	6.5×10^{-1}
glycine	Amino acid	0.01[-0.03: 0.05]	6.5×10^{-1}
serine	Amino acid	0.01[-0.03: 0.05]	6.6×10^{-1}
dimethylarginine (SDMA + ADMA)	Amino acid	0.01[-0.03: 0.05]	6.8×10^{-1}
hypoxanthine	Nucleotide	-0.01[-0.05: 0.03]	6.9×10^{-1}
theobromine	Xenobiotics	-0.01[-0.05: 0.03]	6.9×10^{-1}
erythronate	Carbohydrate	0.01[-0.03: 0.05]	6.9×10^{-1}
taurocholate	Lipid	0.01[-0.04: 0.06]	6.9×10^{-1}
alpha-ketoglutarate	Energy	0.01[-0.03: 0.05]	7.1×10^{-1}
stearidonate (18:4n3)	Lipid	-0.01[-0.04: 0.03]	7.1×10^{-1}
2-palmitoylglycerophosphocholine	Lipid	-0.01[-0.04: 0.03]	7.3×10^{-1}
eicosenoate (20:1n9 or 11)	Lipid	0.01[-0.03: 0.04]	7.4×10^{-1}
pyruvate	Carbohydrate	0.01[-0.03: 0.04]	7.5×10^{-1}
valerate	Lipid	-0.01[-0.06: 0.04]	7.6×10^{-1}
fructose	Carbohydrate	0.01[-0.03: 0.04]	7.6×10^{-1}
EPA	Lipid	0.01[-0.03: 0.04]	7.6×10^{-1}
stearate (18:0)	Lipid	-0.01[-0.04: 0.03]	7.6×10^{-1}
linolenate [alpha or gamma; (18:3n3 or 6)]	Lipid	-0.01[-0.04: 0.03]	7.7×10^{-1}
octadecanedioate	Lipid	0.01[-0.03: 0.04]	7.7×10^{-1}
urea	Amino acid	-0.01[-0.04: 0.03]	7.7×10^{-1}
indoleacetate	Amino acid	0.01[-0.03: 0.04]	7.8×10^{-1}
cortisone	Lipid	-0.01[-0.04: 0.03]	7.8×10^{-1}
levulinate (4-oxovalerate)	Amino acid	0.01[-0.03: 0.04]	7.8×10^{-1}
oleate (18:1n9)	Lipid	-0.01[-0.04: 0.03]	7.8×10^{-1}
margarate (17:0)	Lipid	-0.01[-0.04: 0.03]	7.8×10^{-1}
1-oleoylglycerophosphocholine	Lipid	0.01[-0.03: 0.04]	7.9×10^{-1}
taurothicholate 3-sulfate	Lipid	-0.00[-0.05: 0.04]	8.1×10^{-1}
leucylleucine	Peptide	-0.01[-0.06: 0.05]	8.2×10^{-1}
ADSGEGDFXAEGGGVR	Peptide	0.01[-0.04: 0.05]	8.2×10^{-1}
3-methoxytyrosine	Amino acid	-0.00[-0.05: 0.04]	8.3×10^{-1}
2-hydroxybutyrate (AHB)	Amino acid	0.00[-0.03: 0.04]	8.3×10^{-1}
methionine	Amino acid	0.00[-0.03: 0.04]	8.3×10^{-1}
pipecolate	Amino acid	-0.00[-0.04: 0.03]	8.3×10^{-1}
2-oleoylglycerophosphocholine	Lipid	0.00[-0.03: 0.04]	8.4×10^{-1}
1-linoleoylglycerophosphocholine	Lipid	0.00[-0.03: 0.04]	8.4×10^{-1}
scyllo-inositol	Lipid	0.00[-0.04: 0.05]	8.4×10^{-1}
glutamine	Amino acid	-0.00[-0.04: 0.03]	8.4×10^{-1}
indolelactate	Amino acid	-0.00[-0.04: 0.03]	8.5×10^{-1}
heme	Cofactors and vitamins	0.00[-0.04: 0.05]	8.6×10^{-1}
2-aminobutyrate	Amino acid	0.00[-0.03: 0.04]	8.6×10^{-1}
10-nonadecenoate (19:1n9)	Lipid	-0.00[-0.04: 0.03]	8.8×10^{-1}
1-myristoylglycerophosphocholine	Lipid	-0.00[-0.04: 0.03]	8.8×10^{-1}
thymol sulfate	Xenobiotics	-0.00[-0.06: 0.05]	8.8×10^{-1}
pro-hydroxy-pro	Peptide	0.00[-0.04: 0.04]	8.8×10^{-1}
trans-4-hydroxyproline	Amino acid	0.00[-0.04: 0.04]	8.9×10^{-1}
cortisol	Lipid	0.00[-0.04: 0.04]	8.9×10^{-1}
5-oxoproline	Amino acid	0.00[-0.04: 0.04]	8.9×10^{-1}
threitol	Carbohydrate	-0.00[-0.04: 0.03]	8.9×10^{-1}
G3P	Lipid	0.00[-0.03: 0.04]	9.0×10^{-1}
serotonin	Amino acid	-0.00[-0.05: 0.04]	9.0×10^{-1}
acetylcarnitine	Lipid	-0.00[-0.04: 0.04]	9.0×10^{-1}
isobutyrylcarnitine	Amino acid	0.00[-0.04: 0.04]	9.0×10^{-1}
HWESASXX	Peptide	-0.00[-0.04: 0.03]	9.1×10^{-1}
bilirubin (E,Z or Z,E)	Cofactors and vitamins	0.00[-0.04: 0.05]	9.1×10^{-1}
metoprolol acid metabolite	Xenobiotics	-0.01[-0.19: 0.18]	9.2×10^{-1}
1-docosahexaenoylglycerophosphocholine	Lipid	0.00[-0.04: 0.04]	9.4×10^{-1}
ergothioneine	Xenobiotics	-0.00[-0.05: 0.05]	9.4×10^{-1}
p-cresol sulfate	Amino acid	0.00[-0.04: 0.04]	9.4×10^{-1}
indolepropionate	Amino acid	0.00[-0.04: 0.04]	9.4×10^{-1}
choline	Lipid	0.00[-0.04: 0.04]	9.4×10^{-1}
isovalerylcarnitine	Amino acid	-0.00[-0.04: 0.04]	9.5×10^{-1}
2-linoleoylglycerophosphocholine	Lipid	-0.00[-0.04: 0.04]	9.5×10^{-1}
docosapentaenoate (n3 DPA; 22:5n3)	Lipid	0.00[-0.04: 0.04]	9.5×10^{-1}
1-eicosadienoylglycerophosphocholine	Lipid	0.00[-0.04: 0.04]	9.5×10^{-1}
1-palmitoylglycerophosphocholine	Lipid	-0.00[-0.04: 0.04]	9.5×10^{-1}
phosphate	Energy	-0.00[-0.04: 0.04]	9.5×10^{-1}
phenylacetylglutamine	Amino acid	0.00[-0.04: 0.04]	9.6×10^{-1}
3-methylxanthine	Xenobiotics	0.00[-0.04: 0.05]	9.6×10^{-1}
DSGEGDFXAEGGGVR	Peptide	-0.00[-0.05: 0.05]	9.6×10^{-1}
xanthine	Nucleotide	-0.00[-0.04: 0.04]	9.8×10^{-1}
gamma-glutamylglutamine	Peptide	-0.00[-0.04: 0.04]	10.0×10^{-1}

Figure A.1 Meta-analysis of metabolites associated with telomere length. Five metabolites were significantly correlated with leukocyte telomere length (LTL) in the TwinsUK cohort and subsequently replicated in the KORA cohort. Results from both cohorts were meta-analysed using fixed-effects meta-analysis.

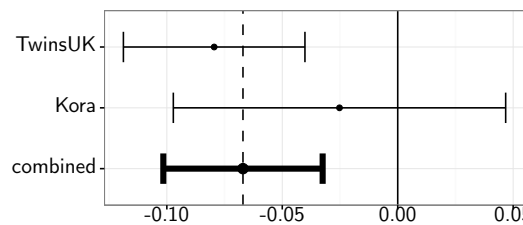
(a) 1-palmitoylglycerophosphoinositol



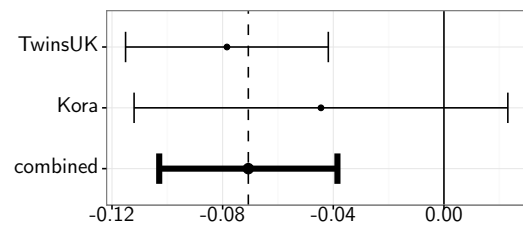
(b) 1-stearoylglycerophosphoinositol



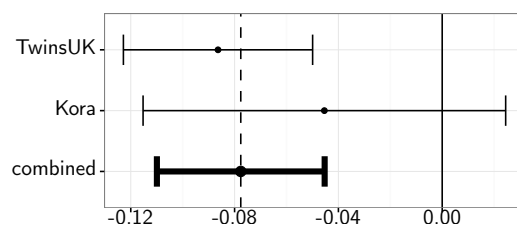
(c) 4-vinylphenol sulfate



(d) gamma-glutamylphenylalanine



(e) gamma-glutamyltyrosine



APPENDIX B

Glycosylation profile of IgG in moderate kidney dysfunction

Table B.1 Associations of IgG glycan traits with renal function. 76 glycan traits (as described in Table 3.3) were tested for association with renal function in 3212 individuals of the discovery cohort. Bonferroni-significant associations are indicated by ***, FDR-significant associations by ** and nominally significant associations by *.

ID	Glycan Trait	β [95%CI]	p-value
GP18	GP18	1.48 [0.89: 2.07]	$8.6 \times 10^{-7***}$
GP14	GP14	1.46 [0.85: 2.07]	$2.9 \times 10^{-6***}$
IGP45	GP6 ⁿ	-1.39 [-1.98: -0.80]	$3.6 \times 10^{-6***}$
IGP53	GP14 ⁿ	1.29 [0.68: 1.90]	$3.1 \times 10^{-5***}$
IGP37	FBS1 / FS1	-1.12 [-1.65: -0.59]	$3.5 \times 10^{-5***}$
IGP38	FBS1 / (FS1+FBS1)	-1.10 [-1.63: -0.57]	$4.6 \times 10^{-5***}$
IGP36	FBS ^{total} / FS ^{total}	-1.07 [-1.60: -0.54]	$8.2 \times 10^{-5***}$
IGP57	G2 ⁿ	1.20 [0.60: 1.80]	$8.8 \times 10^{-5***}$
GP6	GP6	-1.14 [-1.71: -0.57]	$8.9 \times 10^{-5***}$
IGP55	G0 ⁿ	-1.16 [-1.76: -0.56]	$1.5 \times 10^{-4***}$
IGP26	FGS / (F + FG + FGS)	1.01 [0.46: 1.56]	$3.0 \times 10^{-4***}$
IGP42	GP2 ⁿ	-0.91 [-1.42: -0.40]	$5.0 \times 10^{-4***}$
GP2	GP2	-0.90 [-1.42: -0.38]	$6.3 \times 10^{-4***}$
IGP77	BG2 ⁿ / (FG2 ⁿ + FBG2 ⁿ)	-0.93 [-1.46: -0.39]	$6.6 \times 10^{-4***}$
IGP76	FG2 ⁿ / (BG2 ⁿ + FBG2 ⁿ)	0.91 [0.38: 1.44]	$7.3 \times 10^{-4**}$
IGP39	FBS2 / FS2	-0.90 [-1.42: -0.37]	$8.5 \times 10^{-4**}$
IGP40	FBS2 / (FS2+FBS2)	-0.91 [-1.44: -0.37]	$8.5 \times 10^{-4**}$
IGP75	FBG2 ⁿ / (FG2 ⁿ + FBG2 ⁿ)	-0.89 [-1.42: -0.37]	$8.5 \times 10^{-4**}$
IGP74	FBG2 ⁿ / FG2 ⁿ	-0.89 [-1.41: -0.36]	$9.0 \times 10^{-4**}$
IGP69	FBG2 ⁿ / G2 ⁿ	-0.85 [-1.37: -0.33]	$1.4 \times 10^{-3**}$
IGP43	GP4 ⁿ	-0.90 [-1.47: -0.33]	$2.0 \times 10^{-3**}$
IGP24	FGS / (FG + FGS)	0.76 [0.25: 1.27]	$3.4 \times 10^{-3**}$
IGP65	FG2 ⁿ / G2 ⁿ	0.73 [0.23: 1.24]	$4.4 \times 10^{-3**}$
GP23	GP23	0.69 [0.19: 1.18]	$6.6 \times 10^{-3**}$
IGP71	FB ⁿ / F ⁿ total	-0.69 [-1.19: -0.18]	$7.8 \times 10^{-3**}$
IGP66	FB ⁿ	-0.68 [-1.19: -0.17]	$8.5 \times 10^{-3**}$
IGP70	FB ⁿ / F ⁿ	-0.67 [-1.17: -0.16]	$9.4 \times 10^{-3**}$
GP15	GP15	0.70 [0.17: 1.23]	$9.5 \times 10^{-3**}$
IGP72	F ⁿ / (B ⁿ + FB ⁿ)	0.65 [0.15: 1.16]	$1.1 \times 10^{-2**}$
IGP31	FBG2S1 / (FBG2 + FBG2S1 + FBG2S2)	-0.63 [-1.12: -0.13]	$1.3 \times 10^{-2**}$
IGP62	F ⁿ	0.58 [0.09: 1.07]	$2.2 \times 10^{-2*}$
GP7	GP7	-0.53 [-1.00: -0.06]	$2.8 \times 10^{-2*}$
IGP46	GP7 ⁿ	-0.52 [-0.99: -0.05]	$2.9 \times 10^{-2*}$
GP4	GP4	-0.60 [-1.14: -0.06]	$2.9 \times 10^{-2*}$
IGP60	FG1 ⁿ total / G1 ⁿ	0.53 [0.05: 1.00]	$2.9 \times 10^{-2*}$
GP16	GP16	0.51 [0.03: 0.99]	$3.7 \times 10^{-2*}$
IGP47	GP8 ⁿ	0.53 [0.02: 1.04]	$4.0 \times 10^{-2*}$
IGP59	FG0 ⁿ total / G0 ⁿ	0.51 [0.02: 1.00]	$4.0 \times 10^{-2*}$
IGP50	GP11 ⁿ	-0.50 [-0.99: -0.01]	$4.7 \times 10^{-2*}$
IGP33	F ^{total} S1 / F ^{total} S2	0.49 [0.00: 0.98]	$4.8 \times 10^{-2*}$
IGP54	GP15 ⁿ	0.51 [-0.01: 1.03]	5.5×10^{-2}
GP8	GP8	0.46 [-0.02: 0.93]	5.9×10^{-2}
IGP61	FG2 ⁿ total / G2 ⁿ	0.44 [-0.04: 0.92]	7.5×10^{-2}
IGP67	FBG0 ⁿ / G0 ⁿ	-0.42 [-0.90: 0.05]	7.9×10^{-2}
IGP63	FG0 ⁿ / G0 ⁿ	0.36 [-0.11: 0.83]	1.3×10^{-1}
GP11	GP11	-0.38 [-0.88: 0.11]	1.3×10^{-1}
IGP64	FG1 ⁿ / G1 ⁿ	0.36 [-0.12: 0.85]	1.4×10^{-1}
IGP68	FBG1 ⁿ / G1 ⁿ	-0.36 [-0.85: 0.12]	1.4×10^{-1}
IGP56	G1 ⁿ	0.36 [-0.12: 0.84]	1.4×10^{-1}
GP13	GP13	0.33 [-0.16: 0.82]	1.9×10^{-1}

continued on next page ...

IGP41	GP1 ⁿ	-0.32 [-0.81: 0.16]	1.9×10^{-1}
IGP58	F ⁿ total	0.31 [-0.16: 0.77]	1.9×10^{-1}
IGP35	FBS1 / FBS2	-0.30 [-0.78: 0.17]	2.1×10^{-1}
IGP30	FG2S2 / (FG2 + FG2S1 + FG2S2)	-0.27 [-0.75: 0.21]	2.7×10^{-1}
IGP44	GP5 ⁿ	-0.26 [-0.75: 0.22]	2.9×10^{-1}
GP1	GP1	-0.26 [-0.74: 0.22]	2.9×10^{-1}
IGP29	FG2S1 / (FG2 + FG2S1 + FG2S2)	0.26 [-0.24: 0.76]	3.1×10^{-1}
IGP27	FBGS / (FB + FBG + FBGS)	0.24 [-0.23: 0.72]	3.2×10^{-1}
GP5	GP5	-0.24 [-0.73: 0.25]	3.3×10^{-1}
GP2021	GP20 + GP21	0.14 [-0.14: 0.42]	3.4×10^{-1}
IGP49	GP10 ⁿ	-0.21 [-0.68: 0.26]	3.8×10^{-1}
IGP28	FG1S1 / (FG1 + FG1S1)	0.20 [-0.28: 0.67]	4.2×10^{-1}
IGP52	GP13 ⁿ	0.20 [-0.29: 0.69]	4.2×10^{-1}
GP12	GP12	0.18 [-0.30: 0.66]	4.6×10^{-1}
IGP73	B ⁿ / (F ⁿ + FB ⁿ)	0.18 [-0.31: 0.66]	4.7×10^{-1}
IGP34	FS1 / FS2	0.17 [-0.30: 0.64]	4.7×10^{-1}
GP22	GP22	0.13 [-0.32: 0.58]	5.7×10^{-1}
GP10	GP10	-0.13 [-0.60: 0.34]	5.9×10^{-1}
IGP51	GP12 ⁿ	0.13 [-0.35: 0.61]	5.9×10^{-1}
GP19	GP19	-0.11 [-0.57: 0.35]	6.4×10^{-1}
IGP32	FBG2S2 / (FBG2 + FBG2S1 + FBG2S2)	-0.11 [-0.59: 0.37]	6.6×10^{-1}
GP17	GP17	-0.09 [-0.55: 0.37]	7.1×10^{-1}
GP24	GP24	0.08 [-0.38: 0.53]	7.4×10^{-1}
GP9	GP9	0.07 [-0.39: 0.53]	7.7×10^{-1}
IGP25	FBGS / (FBG + FBGS)	-0.04 [-0.51: 0.42]	8.5×10^{-1}
IGP48	GP9 ⁿ	-0.02 [-0.48: 0.45]	9.5×10^{-1}

Table B.2 Associations of whole plasma glycan traits with renal function. To test whether observed associations of glycosylation and renal function are specific to IgG or reflect a more global pattern of protein glycosylation, I correlated whole plasma glycans with eGFR. None of the associations passed multiple-testing correction, nominally significant associations are indicated by *.

ID	Main Glycan	β [95%CI]	p-value
gly2	M5, FA2B	-1.91 [-3.42: -0.40]	$1.3 \times 10^{-2*}$
gly1	FA2	-1.44 [-2.88: -0.00]	$4.9 \times 10^{-2*}$
gly42	A4F1G4S4	0.89 [-0.27: 2.06]	1.3×10^{-1}
gly34	FA3G3S3	0.88 [-0.33: 2.09]	1.5×10^{-1}
gly41	A4G4S4	0.84 [-0.31: 2.00]	1.5×10^{-1}
gly6	FA2[6]BG1	-0.96 [-2.32: 0.40]	1.7×10^{-1}
gly30	A3G3S3	0.76 [-0.53: 2.04]	2.5×10^{-1}
gly16	A2BG2S1	-0.71 [-2.00: 0.58]	2.8×10^{-1}
gly40	A4G4S4	0.64 [-0.55: 1.84]	2.9×10^{-1}
gly4	FA2[6]G1	-0.74 [-2.11: 0.63]	2.9×10^{-1}
gly39	A4G4S4	0.63 [-0.61: 1.87]	3.2×10^{-1}
gly13	FA2[3]G1S1	0.60 [-0.58: 1.77]	3.2×10^{-1}
gly35	A3F1G3S3	0.67 [-0.66: 2.00]	3.2×10^{-1}
gly33	A3G3S3	0.58 [-0.62: 1.78]	3.5×10^{-1}
gly10.11	FA2G2	0.76 [-0.92: 2.43]	3.8×10^{-1}
gly36	A4G4S3	0.55 [-0.71: 1.81]	3.9×10^{-1}
gly3	A2[6]BG1	-0.53 [-1.82: 0.77]	4.3×10^{-1}
gly19	FA2BG2S1	-0.53 [-1.84: 0.78]	4.3×10^{-1}
gly9	A2BG2	-0.49 [-1.70: 0.72]	4.3×10^{-1}
gly14	FA2[3]G1S1	-0.49 [-1.78: 0.81]	4.6×10^{-1}
gly7	M6	-0.44 [-1.61: 0.73]	4.6×10^{-1}
gly38	A4F1G3S3	0.42 [-0.89: 1.73]	5.3×10^{-1}
gly37	A4G4S3	0.40 [-0.90: 1.69]	5.5×10^{-1}
gly18	FA2G2S1	0.44 [-1.07: 1.94]	5.7×10^{-1}
gly26	FA2BG2S2	-0.36 [-1.70: 0.99]	6.0×10^{-1}
gly24	A2BG2S2	0.28 [-0.99: 1.55]	6.7×10^{-1}
gly5	FA2[3]G1	-0.29 [-1.64: 1.05]	6.7×10^{-1}
gly25	FA2G2S2	-0.26 [-1.62: 1.11]	7.1×10^{-1}
gly29	A3G3S2	0.23 [-1.13: 1.59]	7.4×10^{-1}
gly31.32	A3G3S3, FA3G3S3	0.20 [-1.01: 1.40]	7.5×10^{-1}
gly17	M5A1G1S1	-0.19 [-1.52: 1.14]	7.8×10^{-1}
gly27.28	A3G3S2, A3BG3S2	0.16 [-1.07: 1.40]	7.9×10^{-1}
gly22	M9	-0.16 [-1.48: 1.16]	8.1×10^{-1}
gly23	A2G2S2	-0.10 [-1.38: 1.17]	8.7×10^{-1}
gly12	FA2BG2	-0.07 [-1.46: 1.32]	9.2×10^{-1}
gly8	A2G2	0.06 [-1.27: 1.39]	9.3×10^{-1}
gly20.21	A2G2S2	-0.04 [-1.30: 1.21]	9.5×10^{-1}
gly15	A2G2S1	0.02 [-1.12: 1.16]	9.7×10^{-1}

Table B.3 Comparison of renal-associated IgG glycosylation profiles between UPLC and MS/MS measurements. The associations of renal function with *immunoglobulin G* (IgG) glycosylation were compared between glycans measured by ultra-performance liquid chromatography (UPLC) and glycans measured by tandem mass spectrometry (MS/MS) to confirm F γ glycosylation as driving force of the observed associations in a subset of 96 age-matched individuals.

UPLC glycans		MS/MS glycans	
Glycan	β [95 %CI]	Glycan	β [95 %CI]
GP18	1.48 [0.89: 2.07]	IgG1 G2FS1	3.27 [-2.28: 8.82]
GP14	1.46 [0.85: 2.07]	IgG1 G2F	6.23 [-0.59: 13.04]
GP6 ⁿ	-1.39 [-1.98:-0.80]	IgG1 G0FNn	-1.46 [-6.98: 4.07]
GP14 ⁿ	1.29 [0.68: 1.90]	IgG1 G2Fn	5.20 [-1.39: 11.79]
FBS1 / FS1	-1.12 [-1.65:-0.59]	IgG1 FBS1/FS	-4.55 [-9.49: 0.40]
FBS1 / (FS1+FBS1)	-1.10 [-1.63:-0.57]	IgG1 FBS1/(FS1+FBS1)	-5.34 [-10.45: -0.23]
G2 ⁿ	1.20 [0.60: 1.80]	IgG1 G2n	3.44 [-3.04: 9.92]
GP6	-1.14 [-1.71:-0.57]	IgG1 G0FN	-1.15 [-6.73: 4.43]
G0 ⁿ	-1.16 [-1.76:-0.56]	IgG1 G0n	-0.71 [-6.81: 5.38]
FGS / (F + FG + FGS)	1.01 [0.46: 1.56]	IgG1 FGS1/(F+FG+FGS1)	-0.23 [-5.38: 4.91]
GP2 ⁿ	-0.91 [-1.42:-0.40]	IgG1 G0n	-2.27 [-8.33: 3.80]
GP2	-0.90 [-1.43:-0.38]	IgG1 G0	-1.88 [-7.97: 4.21]

APPENDIX C

Metabolic markers of renal disease in type 2 diabetics compared to non-diabetics

Figure C.1 Absolute concentrations of cholesterol. (a) The absolute concentrations of total cholesterol in serum, as well as cholesterol in IDL, LDL, and HDL were all higher in non-diabetic participants of TwinsUK than in diabetic participants of GenoDiabMar. This is most probably due to treatment of dyslipidaemia in diabetics. (b) In contrast, the proportion of triglycerides in various cholesterol subclasses was consistently higher in the GenoDiabMar cohort.

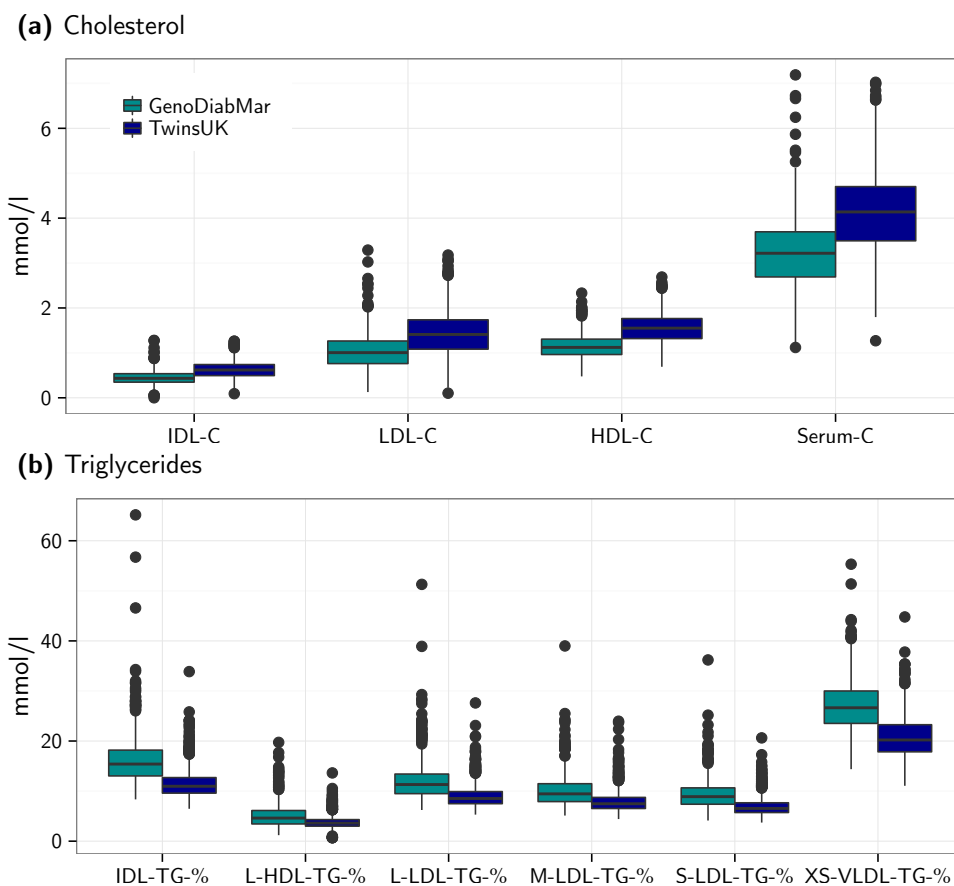


Table C.1 Associations of NMR metabolites with microvascular diseases in the GenoDiab-Mar cohort. Metabolites significantly associated with estimated glomerular filtration rate (eGFR) were tested for association with the concentration of albumin in urine (albuminuria), and diabetic retinopathy.

Metabolite	eGFR		Albuminuria		Diabetic Retinopathy	
	β [95 %CI]	p-value	β [95 %CI]	p-value	OR [95 %CI]	p-value
XXL-VLDL-C	0.10 [-1.98: 2.17]	9.3×10^{-1}	0.04 [-0.20: 0.28]	7.4×10^{-1}	0.88 [0.73:1.07]	2.1×10^{-1}
XXL-VLDL-CE	-0.41 [-2.48: 1.67]	7.0×10^{-1}	0.05 [-0.20: 0.29]	7.2×10^{-1}	0.94 [0.78:1.14]	5.4×10^{-1}
XL-VLDL-C	0.22 [-1.85: 2.30]	8.3×10^{-1}	0.13 [-0.12: 0.37]	3.1×10^{-1}	0.91 [0.75:1.11]	3.6×10^{-1}
XL-VLDL-CE	-0.36 [-2.43: 1.71]	7.4×10^{-1}	0.15 [-0.09: 0.40]	2.2×10^{-1}	0.95 [0.78:1.15]	5.8×10^{-1}
L-VLDL-L	0.06 [-2.02: 2.14]	9.5×10^{-1}	0.15 [-0.09: 0.40]	2.2×10^{-1}	0.91 [0.75:1.11]	3.5×10^{-1}
L-VLDL-PL	-0.14 [-2.22: 1.94]	8.9×10^{-1}	0.18 [-0.07: 0.42]	1.6×10^{-1}	0.93 [0.76:1.12]	4.3×10^{-1}
L-VLDL-C	-0.48 [-2.55: 1.60]	6.5×10^{-1}	0.17 [-0.08: 0.42]	1.8×10^{-1}	0.94 [0.78:1.14]	5.6×10^{-1}
L-VLDL-CE	-0.92 [-3.00: 1.15]	3.8×10^{-1}	0.19 [-0.06: 0.44]	1.3×10^{-1}	0.98 [0.81:1.19]	8.6×10^{-1}
L-VLDL-FC	-0.08 [-2.16: 1.99]	9.4×10^{-1}	0.15 [-0.09: 0.40]	2.3×10^{-1}	0.91 [0.75:1.11]	3.5×10^{-1}
L-VLDL-TG	0.07 [-2.01: 2.15]	9.5×10^{-1}	0.16 [-0.09: 0.40]	2.1×10^{-1}	0.90 [0.75:1.10]	3.1×10^{-1}
M-VLDL-P	-0.96 [-3.03: 1.11]	3.6×10^{-1}	0.20 [-0.05: 0.45]	1.2×10^{-1}	0.95 [0.78:1.15]	6.0×10^{-1}
M-VLDL-L	-0.50 [-2.57: 1.57]	6.3×10^{-1}	0.17 [-0.08: 0.43]	1.7×10^{-1}	0.94 [0.78:1.14]	5.2×10^{-1}
M-VLDL-PL	-0.75 [-2.82: 1.32]	4.8×10^{-1}	0.20 [-0.05: 0.46]	1.2×10^{-1}	0.95 [0.79:1.15]	6.0×10^{-1}
M-VLDL-C	-0.77 [-2.84: 1.30]	4.6×10^{-1}	0.16 [-0.09: 0.41]	2.2×10^{-1}	0.98 [0.81:1.19]	8.5×10^{-1}
M-VLDL-CE	-0.55 [-2.62: 1.52]	6.0×10^{-1}	0.10 [-0.15: 0.35]	4.3×10^{-1}	1.01 [0.83:1.21]	9.5×10^{-1}
M-VLDL-FC	-1.07 [-3.14: 1.00]	3.1×10^{-1}	0.22 [-0.03: 0.47]	8.8×10^{-2}	0.97 [0.80:1.17]	7.2×10^{-1}
M-VLDL-TG	-0.66 [-2.73: 1.41]	5.3×10^{-1}	0.19 [-0.06: 0.44]	1.4×10^{-1}	0.93 [0.77:1.12]	4.4×10^{-1}
S-VLDL-P	-1.92 [-3.97: 0.14]	6.8×10^{-2}	0.31 [0.04: 0.57]	2.3×10^{-2}	1.01 [0.84:1.22]	8.8×10^{-1}
S-VLDL-L	-1.44 [-3.50: 0.62]	1.7×10^{-1}	0.29 [0.03: 0.55]	3.1×10^{-2}	1.00 [0.83:1.20]	9.8×10^{-1}
S-VLDL-PL	-1.78 [-3.84: 0.28]	9.1×10^{-2}	0.32 [0.06: 0.59]	1.8×10^{-2}	1.02 [0.84:1.23]	8.7×10^{-1}
S-VLDL-C	-0.89 [-2.95: 1.17]	4.0×10^{-1}	0.15 [-0.11: 0.40]	2.5×10^{-1}	1.05 [0.87:1.27]	5.9×10^{-1}
S-VLDL-CE	-0.18 [-2.25: 1.88]	8.6×10^{-1}	0.04 [-0.21: 0.28]	7.7×10^{-1}	1.06 [0.88:1.28]	5.4×10^{-1}
S-VLDL-FC	-1.97 [-4.03: 0.09]	6.1×10^{-2}	0.32 [0.06: 0.58]	1.6×10^{-2}	1.05 [0.87:1.26]	6.4×10^{-1}
S-VLDL-TG	-1.99 [-4.04: 0.07]	5.9×10^{-2}	0.35 [0.09: 0.61]	8.8×10^{-3}	0.98 [0.81:1.18]	8.2×10^{-1}
XS-VLDL-P	0.21 [-1.86: 2.28]	8.4×10^{-1}	0.18 [-0.08: 0.44]	1.8×10^{-1}	1.05 [0.87:1.27]	5.9×10^{-1}
XS-VLDL-L	0.79 [-1.28: 2.86]	4.6×10^{-1}	0.14 [-0.12: 0.40]	2.9×10^{-1}	1.03 [0.85:1.24]	7.9×10^{-1}
XS-VLDL-PL	1.60 [-0.46: 3.66]	1.3×10^{-1}	0.09 [-0.16: 0.35]	4.8×10^{-1}	1.06 [0.87:1.28]	5.7×10^{-1}
XS-VLDL-C	2.40 [0.33: 4.46]	2.3×10^{-2}	-0.07 [-0.32: 0.19]	6.2×10^{-1}	0.98 [0.81:1.18]	8.4×10^{-1}
XS-VLDL-FC	0.05 [-2.01: 2.12]	9.6×10^{-1}	0.11 [-0.15: 0.37]	4.0×10^{-1}	1.11 [0.92:1.34]	2.8×10^{-1}
XS-VLDL-TG	-3.13 [-5.18: -1.09]	2.8×10^{-3}	0.50 [0.23: 0.76]	2.4×10^{-4}	1.10 [0.92:1.33]	3.0×10^{-1}
IDL-P	5.25 [3.22: 7.28]	5.2×10^{-7}	-0.10 [-0.35: 0.16]	4.6×10^{-1}	0.91 [0.76:1.10]	3.5×10^{-1}
IDL-L	5.70 [3.68: 7.72]	4.5×10^{-8}	-0.14 [-0.39: 0.12]	3.0×10^{-1}	0.90 [0.74:1.08]	2.6×10^{-1}
IDL-PL	5.35 [3.33: 7.38]	2.8×10^{-7}	-0.12 [-0.38: 0.13]	3.4×10^{-1}	0.94 [0.78:1.13]	5.0×10^{-1}
IDL-C	6.55 [4.55: 8.55]	2.8×10^{-10}	-0.23 [-0.49: 0.02]	7.1×10^{-2}	0.88 [0.73:1.06]	1.8×10^{-1}
IDL-CE	6.90 [4.90: 8.89]	2.9×10^{-11}	-0.25 [-0.51: 0.00]	5.3×10^{-2}	0.85 [0.70:1.03]	9.7×10^{-2}
IDL-FC	5.14 [3.11: 7.17]	8.5×10^{-7}	-0.17 [-0.42: 0.08]	1.8×10^{-1}	0.98 [0.82:1.19]	8.7×10^{-1}
IDL-TG	-3.23 [-5.29: -1.17]	2.2×10^{-3}	0.54 [0.28: 0.80]	5.3×10^{-5}	1.21 [1.01:1.45]	4.2×10^{-2}
L-LDL-P	5.21 [3.19: 7.24]	5.9×10^{-7}	-0.08 [-0.33: 0.17]	5.2×10^{-1}	0.94 [0.78:1.13]	4.9×10^{-1}
L-LDL-L	5.92 [3.91: 7.93]	1.2×10^{-8}	-0.12 [-0.38: 0.14]	3.6×10^{-1}	0.90 [0.74:1.08]	2.5×10^{-1}
L-LDL-PL	6.48 [4.47: 8.48]	4.4×10^{-10}	-0.17 [-0.42: 0.09]	2.0×10^{-1}	0.87 [0.72:1.05]	1.5×10^{-1}
L-LDL-C	6.33 [4.33: 8.33]	9.9×10^{-10}	-0.18 [-0.43: 0.08]	1.8×10^{-1}	0.89 [0.73:1.07]	2.1×10^{-1}
L-LDL-CE	6.15 [4.14: 8.15]	3.0×10^{-9}	-0.17 [-0.42: 0.09]	2.0×10^{-1}	0.89 [0.73:1.07]	2.1×10^{-1}
L-LDL-FC	6.34 [4.34: 8.34]	9.2×10^{-10}	-0.19 [-0.44: 0.07]	1.5×10^{-1}	0.92 [0.76:1.10]	3.6×10^{-1}
L-LDL-TG	-1.96 [-4.03: 0.12]	6.5×10^{-2}	0.47 [0.21: 0.72]	3.7×10^{-4}	1.19 [0.99:1.42]	6.5×10^{-2}
M-LDL-L	5.55 [3.54: 7.56]	9.3×10^{-8}	-0.06 [-0.32: 0.19]	6.3×10^{-1}	0.91 [0.75:1.09]	3.1×10^{-1}
M-LDL-PL	6.74 [4.74: 8.74]	8.3×10^{-11}	-0.09 [-0.35: 0.18]	5.1×10^{-1}	0.82 [0.68:0.99]	4.2×10^{-2}
M-LDL-C	5.56 [3.55: 7.57]	8.4×10^{-8}	-0.11 [-0.36: 0.14]	3.9×10^{-1}	0.92 [0.76:1.11]	3.9×10^{-1}
M-LDL-CE	5.00 [2.98: 7.02]	1.5×10^{-6}	-0.11 [-0.36: 0.14]	4.0×10^{-1}	0.95 [0.79:1.15]	6.1×10^{-1}
M-LDL-FC	7.09 [5.11: 9.08]	6.3×10^{-12}	-0.12 [-0.38: 0.15]	4.0×10^{-1}	0.84 [0.70:1.01]	6.6×10^{-2}
M-LDL-TG	-1.62 [-3.69: 0.46]	1.3×10^{-1}	0.48 [0.22: 0.75]	3.4×10^{-4}	1.18 [0.99:1.42]	7.1×10^{-2}
S-LDL-L	5.73 [3.72: 7.75]	3.6×10^{-8}	-0.08 [-0.34: 0.18]	5.7×10^{-1}	0.90 [0.75:1.08]	2.7×10^{-1}
S-LDL-PL	7.12 [5.12: 9.12]	7.5×10^{-12}	-0.11 [-0.39: 0.16]	4.2×10^{-1}	0.81 [0.67:0.97]	2.2×10^{-2}
S-LDL-C	5.59 [3.58: 7.60]	7.3×10^{-8}	-0.12 [-0.37: 0.14]	3.7×10^{-1}	0.93 [0.77:1.12]	4.2×10^{-1}
S-LDL-CE	5.09 [3.07: 7.11]	1.0×10^{-6}	-0.12 [-0.37: 0.13]	3.7×10^{-1}	0.96 [0.79:1.16]	6.6×10^{-1}
S-LDL-FC	6.77 [4.78: 8.76]	6.1×10^{-11}	-0.11 [-0.37: 0.16]	4.3×10^{-1}	0.84 [0.70:1.01]	5.8×10^{-2}
S-LDL-TG	-1.16 [-3.24: 0.91]	2.7×10^{-1}	0.41 [0.14: 0.67]	2.8×10^{-3}	1.12 [0.93:1.35]	2.2×10^{-1}
XL-HDL-P	2.80 [0.69: 4.91]	9.6×10^{-3}	-0.25 [-0.51: 0.01]	6.0×10^{-2}	1.13 [0.93:1.37]	2.0×10^{-1}
XL-HDL-L	3.44 [1.33: 5.54]	1.4×10^{-3}	-0.26 [-0.53: -0.00]	4.8×10^{-2}	1.11 [0.91:1.34]	3.1×10^{-1}
XL-HDL-PL	2.05 [-0.07: 4.18]	5.9×10^{-2}	-0.26 [-0.52: 0.00]	5.1×10^{-2}	1.18 [0.97:1.43]	9.0×10^{-2}
XL-HDL-C	4.18 [2.10: 6.26]	9.0×10^{-5}	-0.25 [-0.51: 0.00]	5.4×10^{-2}	1.04 [0.86:1.26]	6.8×10^{-1}
XL-HDL-CE	4.22 [2.15: 6.30]	7.4×10^{-5}	-0.23 [-0.49: 0.02]	7.5×10^{-2}	1.04 [0.86:1.26]	6.9×10^{-1}
XL-HDL-FC	3.60 [1.50: 5.70]	8.3×10^{-4}	-0.28 [-0.53: -0.03]	2.7×10^{-2}	1.06 [0.87:1.28]	5.9×10^{-1}
L-HDL-P	1.65 [-0.49: 3.79]	1.3×10^{-1}	-0.32 [-0.57: -0.06]	1.6×10^{-2}	1.14 [0.94:1.38]	1.8×10^{-1}
L-HDL-L	2.21 [0.08: 4.34]	4.3×10^{-2}	-0.34 [-0.59: -0.08]	1.1×10^{-2}	1.11 [0.91:1.35]	2.9×10^{-1}
L-HDL-PL	2.53 [0.40: 4.66]	2.0×10^{-2}	-0.36 [-0.62: -0.11]	5.5×10^{-3}	1.08 [0.89:1.31]	4.2×10^{-1}
L-HDL-C	1.93 [-0.19: 4.05]	7.5×10^{-2}	-0.30 [-0.56: -0.05]	2.1×10^{-2}	1.13 [0.93:1.37]	2.2×10^{-1}
L-HDL-CE	1.99 [-0.13: 4.12]	6.6×10^{-2}	-0.30 [-0.56: -0.05]	2.0×10^{-2}	1.12 [0.92:1.35]	2.6×10^{-1}
L-HDL-FC	1.25 [-0.87: 3.37]	2.5×10^{-1}	-0.27 [-0.53: -0.02]	3.5×10^{-2}	1.18 [0.98:1.43]	8.4×10^{-2}
M-HDL-P	5.37 [3.28: 7.46]	6.3×10^{-7}	-0.43 [-0.68: -0.18]	9.4×10^{-4}	0.76 [0.62:0.93]	6.8×10^{-3}
M-HDL-L	5.42 [3.34: 7.50]	4.4×10^{-7}	-0.45 [-0.70: -0.20]	5.4×10^{-4}	0.75 [0.62:0.92]	4.9×10^{-3}

continued on next page ...

M-HDL-PL	5.09[2.99: 7.19]	2.4×10^{-6}	-0.39[-0.64:-0.13]	2.8×10^{-3}	0.77[0.63:0.94]	1.2×10^{-2}
M-HDL-C	5.92[3.86: 7.98]	2.5×10^{-8}	-0.51[-0.76:-0.26]	8.6×10^{-5}	0.74[0.61:0.90]	2.7×10^{-3}
M-HDL-CE	5.96[3.90: 8.01]	2.0×10^{-8}	-0.51[-0.76:-0.26]	6.8×10^{-5}	0.74[0.61:0.89]	2.0×10^{-3}
M-HDL-FC	5.59[3.51: 7.66]	1.8×10^{-7}	-0.45[-0.71:-0.20]	5.1×10^{-4}	0.78[0.64:0.95]	1.5×10^{-2}
S-HDL-TG	-3.73[-5.78:-1.67]	4.0×10^{-4}	0.46[0.22:0.70]	2.2×10^{-4}	1.06[0.88:1.28]	5.3×10^{-1}
Serum-C	6.36[4.33: 8.39]	1.4×10^{-9}	-0.24[-0.50:0.02]	7.3×10^{-2}	0.87[0.72:1.05]	1.5×10^{-1}
VLDL-C	0.57[-1.50: 2.64]	5.9×10^{-1}	0.07[-0.18:0.33]	5.7×10^{-1}	0.95[0.79:1.14]	5.7×10^{-1}
Remnant-C	3.44[1.38: 5.49]	1.1×10^{-3}	-0.07[-0.32:0.19]	6.1×10^{-1}	0.88[0.73:1.06]	1.8×10^{-1}
LDL-C	6.72[4.73: 8.70]	7.5×10^{-11}	-0.17[-0.44:0.10]	2.2×10^{-1}	0.85[0.71:1.02]	7.8×10^{-2}
HDL-C	5.29[3.21: 7.38]	8.0×10^{-7}	-0.44[-0.70:-0.19]	8.0×10^{-4}	0.90[0.74:1.09]	2.7×10^{-1}
HDL2-C	5.71[3.64: 7.78]	8.9×10^{-8}	-0.47[-0.74:-0.20]	6.2×10^{-4}	0.90[0.75:1.09]	3.0×10^{-1}
EstC	7.35[5.35: 9.34]	1.5×10^{-12}	-0.29[-0.55:-0.02]	3.4×10^{-2}	0.82[0.68:0.99]	3.7×10^{-2}
Serum-TG	-1.10[-3.18: 0.98]	3.0×10^{-1}	0.25[0.01:0.50]	4.6×10^{-2}	0.99[0.82:1.19]	9.1×10^{-1}
VLDL-TG	-0.11[-2.18: 1.97]	9.2×10^{-1}	0.18[-0.07:0.43]	1.6×10^{-1}	0.93[0.77:1.12]	4.4×10^{-1}
LDL-TG	-2.34[-4.42:-0.25]	2.9×10^{-2}	0.46[0.21:0.71]	3.2×10^{-4}	1.23[1.01:1.48]	3.7×10^{-2}
HDL-TG	-3.66[-5.78:-1.55]	7.2×10^{-4}	0.16[-0.09:0.41]	2.1×10^{-1}	1.18[0.97:1.43]	9.4×10^{-2}
TG/PG	-2.70[-4.75:-0.64]	1.0×10^{-2}	0.33[0.08:0.57]	1.0×10^{-2}	0.95[0.79:1.14]	5.8×10^{-1}
ApoA1	5.69[3.58: 7.79]	1.6×10^{-7}	-0.42[-0.67:-0.16]	1.4×10^{-3}	0.87[0.72:1.06]	1.8×10^{-1}
ApoB/ApoA1	0.56[-1.48: 2.61]	5.9×10^{-1}	0.20[-0.05:0.45]	1.2×10^{-1}	0.97[0.80:1.17]	7.3×10^{-1}
MUFA	-0.76[-2.84: 1.32]	4.8×10^{-1}	0.14[-0.11:0.40]	2.6×10^{-1}	0.98[0.81:1.19]	8.6×10^{-1}
Lac	-0.30[-2.37: 1.77]	7.8×10^{-1}	-0.23[-0.51:0.04]	9.4×10^{-2}	1.04[0.86:1.25]	7.1×10^{-1}
Pyr	4.82[2.76: 6.88]	5.5×10^{-6}	-0.35[-0.60:-0.09]	8.5×10^{-3}	0.84[0.70:1.02]	7.3×10^{-2}
Cit	-4.97[-7.01:-2.93]	2.2×10^{-6}	-0.12[-0.37:0.14]	3.6×10^{-1}	1.14[0.95:1.38]	1.6×10^{-1}
GloI	-11.03[-13.96:-8.10]	1.7×10^{-12}	0.00[-0.43:0.43]	$1.0 \times 10^{+0}$	1.32[0.99:1.77]	6.1×10^{-2}
Ala	7.14[5.11: 9.17]	1.2×10^{-11}	-0.10[-0.36:0.16]	4.5×10^{-1}	0.83[0.68:1.00]	5.1×10^{-2}
Gly	-12.25[-14.10:-10.41]	1.4×10^{-34}	0.64[0.36:0.92]	8.5×10^{-6}	1.76[1.44:2.16]	4.3×10^{-8}
Ile	2.20[0.13: 4.27]	3.7×10^{-2}	0.03[-0.22:0.28]	8.1×10^{-1}	0.83[0.69:1.00]	5.3×10^{-2}
Leu	5.37[3.34: 7.41]	3.0×10^{-7}	-0.02[-0.28:0.25]	9.1×10^{-1}	0.76[0.63:0.91]	3.5×10^{-3}
Val	8.79[6.83: 10.75]	1.2×10^{-17}	-0.15[-0.44:0.14]	3.0×10^{-1}	0.74[0.62:0.89]	1.6×10^{-3}
Phe	-11.70[-13.54:-9.86]	4.4×10^{-32}	0.44[0.18:0.71]	1.3×10^{-3}	1.32[1.08:1.60]	5.3×10^{-3}
Alb	6.57[4.55: 8.58]	3.3×10^{-10}	-0.67[-0.95:-0.40]	2.5×10^{-6}	0.69[0.57:0.84]	1.4×10^{-4}
Gp	-3.63[-5.72:-1.53]	7.4×10^{-4}	0.15[-0.10:0.41]	2.4×10^{-1}	0.88[0.73:1.07]	2.1×10^{-1}
XXL-VLDL-C-%	-5.33[-7.59:-3.07]	4.9×10^{-6}	0.12[-0.14:0.39]	3.7×10^{-1}	1.21[0.97:1.50]	9.1×10^{-2}
XXL-VLDL-FC-%	-1.17[-3.45: 1.11]	3.2×10^{-1}	0.06[-0.21:0.32]	6.7×10^{-1}	1.02[0.81:1.28]	8.7×10^{-1}
XXL-VLDL-TG-%	3.82[1.56: 6.08]	1.0×10^{-3}	-0.21[-0.53:0.10]	1.9×10^{-1}	0.98[0.80:1.22]	8.8×10^{-1}
M-VLDL-FC-%	-0.64[-2.74: 1.46]	5.5×10^{-1}	0.34[0.08:0.59]	1.1×10^{-2}	0.98[0.81:1.17]	7.9×10^{-1}
XS-VLDL-PL-%	4.40[2.37: 6.42]	2.4×10^{-5}	-0.14[-0.39:0.12]	3.0×10^{-1}	1.06[0.88:1.27]	5.5×10^{-1}
XS-VLDL-C-%	5.30[3.28: 7.31]	3.3×10^{-7}	-0.64[-0.88:-0.39]	4.3×10^{-7}	0.83[0.69:0.99]	4.4×10^{-2}
XS-VLDL-CE-%	6.15[4.16: 8.15]	2.5×10^{-9}	-0.70[-0.95:-0.46]	4.0×10^{-8}	0.76[0.63:0.91]	2.8×10^{-3}
XS-VLDL-TG-%	-5.83[-7.83:-3.83]	1.7×10^{-8}	0.54[0.30:0.79]	1.9×10^{-5}	1.11[0.92:1.34]	2.7×10^{-1}
IDL-PL-%	-2.10[-4.19:-0.02]	4.8×10^{-2}	0.01[-0.24:0.26]	9.5×10^{-1}	1.12[0.93:1.35]	2.4×10^{-1}
IDL-C-%	8.96[7.04: 10.89]	9.1×10^{-19}	-0.66[-0.93:-0.39]	2.3×10^{-6}	0.75[0.63:0.89]	1.4×10^{-3}
IDL-CE-%	9.23[7.31: 11.15]	7.6×10^{-20}	-0.69[-0.94:-0.43]	1.6×10^{-7}	0.64[0.53:0.78]	6.6×10^{-6}
IDL-TG-%	-9.26[-11.18:-7.34]	6.2×10^{-20}	0.74[0.48:0.99]	3.4×10^{-8}	1.32[1.10:1.59]	2.8×10^{-3}
L-LDL-C-%	7.73[5.77: 9.69]	4.0×10^{-14}	-0.50[-0.80:-0.21]	8.5×10^{-4}	0.84[0.71:1.00]	5.6×10^{-2}
L-LDL-CE-%	7.11[5.14: 9.09]	4.2×10^{-12}	-0.42[-0.72:-0.12]	6.8×10^{-3}	0.84[0.70:0.99]	4.1×10^{-2}
L-LDL-FC-%	1.54[-0.54: 3.62]	1.5×10^{-1}	-0.27[-0.53:-0.01]	4.4×10^{-2}	1.02[0.84:1.24]	8.4×10^{-1}
L-LDL-TG-%	-9.26[-11.19:-7.33]	8.9×10^{-20}	0.68[0.42:0.94]	5.0×10^{-7}	1.36[1.13:1.63]	9.1×10^{-4}
M-LDL-C-%	5.09[3.09: 7.10]	8.0×10^{-7}	-0.29[-0.56:-0.02]	3.6×10^{-2}	0.95[0.79:1.14]	5.7×10^{-1}
M-LDL-CE-%	3.49[1.48: 5.51]	7.2×10^{-4}	-0.21[-0.46:0.05]	1.1×10^{-1}	0.97[0.81:1.17]	7.6×10^{-1}
M-LDL-TG-%	-8.87[-10.81:-6.93]	3.7×10^{-18}	0.64[0.38:0.90]	2.2×10^{-6}	1.36[1.13:1.64]	1.3×10^{-3}
S-LDL-C-%	5.18[3.17: 7.18]	5.5×10^{-7}	-0.22[-0.49:0.05]	1.1×10^{-1}	0.94[0.78:1.13]	5.1×10^{-1}
S-LDL-CE-%	3.44[1.42: 5.46]	9.0×10^{-4}	-0.17[-0.43:0.08]	1.8×10^{-1}	1.01[0.84:1.21]	9.5×10^{-1}
S-LDL-TG-%	-8.13[-10.10:-6.17]	2.5×10^{-15}	0.57[0.30:0.83]	2.8×10^{-5}	1.29[1.07:1.56]	7.0×10^{-3}
XL-HDL-PL-%	-2.14[-4.31: 0.03]	5.3×10^{-2}	-0.09[-0.34:0.16]	4.8×10^{-1}	1.27[1.01:1.60]	4.2×10^{-2}
XL-HDL-C-%	3.19[1.05: 5.34]	3.6×10^{-3}	0.05[-0.21:0.32]	7.0×10^{-1}	0.81[0.66:0.99]	3.7×10^{-2}
XL-HDL-CE-%	2.30[0.13: 4.46]	3.8×10^{-2}	0.14[-0.12:0.41]	2.9×10^{-1}	0.82[0.67:1.01]	6.4×10^{-2}
XL-HDL-TG-%	-0.39[-2.53: 1.74]	7.2×10^{-1}	0.15[-0.10:0.39]	2.4×10^{-1}	0.93[0.77:1.14]	5.1×10^{-1}
L-HDL-PL-%	0.84[-1.26: 2.94]	4.3×10^{-1}	0.03[-0.24:0.31]	8.2×10^{-1}	0.78[0.64:0.95]	1.6×10^{-2}
L-HDL-C-%	2.81[0.74: 4.87]	7.9×10^{-3}	-0.19[-0.46:0.09]	1.9×10^{-1}	1.04[0.86:1.26]	6.9×10^{-1}
L-HDL-CE-%	2.92[0.87: 4.98]	5.4×10^{-3}	-0.14[-0.42:0.13]	3.1×10^{-1}	0.98[0.81:1.18]	8.0×10^{-1}
L-HDL-FC-%	2.69[0.62: 4.77]	1.1×10^{-2}	-0.25[-0.53:0.03]	7.7×10^{-2}	1.14[0.93:1.40]	1.9×10^{-1}
L-HDL-TG-%	-6.58[-8.60:-4.56]	3.6×10^{-10}	0.23[-0.02:0.48]	6.9×10^{-2}	1.32[1.09:1.60]	4.5×10^{-3}
M-HDL-PL-%	-5.05[-7.08:-3.02]	1.3×10^{-6}	0.63[0.37:0.88]	2.0×10^{-6}	1.36[1.14:1.63]	7.9×10^{-4}
M-HDL-C-%	5.44[3.41: 7.47]	2.0×10^{-7}	-0.62[-0.87:-0.36]	3.0×10^{-6}	0.74[0.62:0.88]	1.0×10^{-3}
M-HDL-CE-%	4.90[2.85: 6.94]	3.3×10^{-6}	-0.58[-0.83:-0.32]	1.0×10^{-5}	0.74[0.61:0.88]	9.1×10^{-4}
M-HDL-FC-%	5.84[3.80: 7.87]	2.8×10^{-8}	-0.49[-0.77:-0.22]	4.0×10^{-4}	0.82[0.68:0.98]	3.2×10^{-2}
M-HDL-TG-%	-4.75[-6.80:-2.71]	6.2×10^{-6}	0.44[0.18:0.70]	1.0×10^{-3}	1.19[0.99:1.44]	6.3×10^{-2}
S-HDL-TG-%	-5.59[-7.61:-3.57]	8.6×10^{-8}	0.51[0.26:0.75]	5.1×10^{-5}	1.22[1.02:1.47]	3.4×10^{-2}
MUFA/FA	-5.72[-7.73:-3.70]	4.0×10^{-8}	0.48[0.23:0.73]	2.2×10^{-4}	1.17[0.97:1.42]	1.1×10^{-1}
SFA/FA	4.57[2.55: 6.59]	1.1×10^{-5}	-0.25[-0.50:-0.00]	4.7×10^{-2}	0.77[0.63:0.93]	7.7×10^{-3}
HDL-D	2.02[-0.12: 4.15]	6.4×10^{-2}	-0.38[-0.64:-0.12]	4.5×10^{-3}	1.11[0.92:1.35]	2.7×10^{-1}

Table C.2 Longitudinal associations of metabolites with eGFR. Longitudinal follow-ups of estimated glomerular filtration rate (eGFR) and NMR metabolite measurements were available for TwinsUK and YoungFinns. I regressed longitudinal trajectories of renal function against the trajectories of metabolite levels. In contrast, only eGFR follow ups were available for KORA and GenoDiabMar. For these cohorts, I predicted the longitudinal change of eGFR based on baseline metabolite concentrations, correcting for baseline eGFR.

Metabolite	TwinsUK		YoungFinns		GenoDiabMar		KORA	
	β [95 %CI]	p-value	β [95 %CI]	p-value	β [95 %CI]	p-value	β [95 %CI]	p-value
Cit	-0.04 [-0.06;-0.02]	1.5×10^{-6}			0.22 [-0.01;0.45]	6.3×10^{-2}	0.01 [-0.02; 0.04]	7.0×10^{-1}
XL-HDL-PL	0.04 [0.02; 0.05]	1.3×10^{-5}	0.00 [-0.01; 0.01]	4.7×10^{-1}	0.08 [-0.16;0.31]	5.4×10^{-1}	0.02 [-0.01; 0.05]	2.2×10^{-1}
XL-HDL-L	0.04 [0.02; 0.05]	2.2×10^{-5}	0.01 [-0.00; 0.02]	2.1×10^{-1}	0.06 [-0.17;0.29]	6.1×10^{-1}	0.02 [-0.02; 0.05]	3.2×10^{-1}
HDL2-C	0.04 [0.02; 0.05]	4.2×10^{-5}	-0.00 [-0.01; 0.01]	6.1×10^{-1}	0.12 [-0.12;0.36]	3.4×10^{-1}	0.02 [-0.01; 0.05]	2.5×10^{-1}
XL-HDL-P	0.03 [0.02; 0.05]	6.9×10^{-5}	0.01 [-0.00; 0.02]	1.8×10^{-1}	0.07 [-0.16;0.31]	5.3×10^{-1}	0.01 [-0.02; 0.05]	3.8×10^{-1}
L-HDL-L	0.03 [0.02; 0.05]	9.0×10^{-5}	0.01 [-0.00; 0.02]	2.9×10^{-1}	0.15 [-0.09;0.39]	2.2×10^{-1}	0.01 [-0.02; 0.05]	3.9×10^{-1}
L-HDL-C	0.03 [0.02; 0.05]	9.9×10^{-5}	0.01 [-0.00; 0.01]	3.2×10^{-1}	0.14 [-0.09;0.38]	2.4×10^{-1}	0.01 [-0.02; 0.05]	4.0×10^{-1}
L-HDL-CE	0.03 [0.02; 0.05]	9.9×10^{-5}	0.01 [-0.00; 0.01]	3.2×10^{-1}	0.14 [-0.09;0.38]	2.3×10^{-1}	0.01 [-0.02; 0.04]	4.3×10^{-1}
L-HDL-PL	0.03 [0.02; 0.05]	1.0×10^{-4}	0.00 [-0.01; 0.01]	4.8×10^{-1}	0.16 [-0.08;0.40]	2.0×10^{-1}	0.01 [-0.02; 0.05]	3.6×10^{-1}
HDL-C	0.03 [0.02; 0.05]	1.1×10^{-4}	-0.00 [-0.01; 0.01]	6.4×10^{-1}	0.11 [-0.13;0.35]	3.8×10^{-1}	0.01 [-0.02; 0.04]	4.2×10^{-1}
XL-HDL-C	0.03 [0.02; 0.05]	1.2×10^{-4}	0.01 [-0.00; 0.02]	2.7×10^{-1}	0.05 [-0.18;0.28]	6.5×10^{-1}	0.01 [-0.02; 0.04]	5.6×10^{-1}
XL-HDL-FC	0.03 [0.02; 0.05]	1.3×10^{-4}	0.00 [-0.01; 0.01]	4.2×10^{-1}	0.07 [-0.17;0.30]	5.8×10^{-1}		
L-HDL-P	0.03 [0.02; 0.05]	1.6×10^{-4}	0.00 [-0.01; 0.01]	4.5×10^{-1}	0.16 [-0.08;0.40]	2.0×10^{-1}	0.01 [-0.02; 0.05]	4.0×10^{-1}
XL-HDL-CE	0.03 [0.02; 0.05]	1.9×10^{-4}	0.01 [-0.00; 0.02]	2.6×10^{-1}	0.05 [-0.18;0.28]	6.5×10^{-1}	0.01 [-0.02; 0.04]	7.4×10^{-1}
HDL-D	0.03 [0.01; 0.05]	2.0×10^{-4}	0.01 [-0.00; 0.02]	8.7×10^{-2}	0.13 [-0.11;0.36]	2.9×10^{-1}	0.02 [-0.01; 0.05]	2.2×10^{-1}
L-HDL-FC	0.03 [0.02; 0.05]	2.2×10^{-4}	0.00 [-0.01; 0.01]	4.6×10^{-1}	0.14 [-0.10;0.37]	2.5×10^{-1}	0.01 [-0.02; 0.05]	3.6×10^{-1}
L-HDL-C-%	0.03 [0.01; 0.05]	4.9×10^{-4}	0.01 [-0.00; 0.02]	2.3×10^{-1}	0.08 [-0.13;0.29]	4.6×10^{-1}	0.02 [-0.01; 0.05]	2.8×10^{-1}
M-HDL-CE-%	0.03 [0.01; 0.04]	1.4×10^{-3}	-0.00 [-0.01; 0.01]	3.4×10^{-1}	0.12 [-0.11;0.34]	3.1×10^{-1}	0.01 [-0.02; 0.04]	4.2×10^{-1}
L-HDL-PL-%	-0.03 [-0.04;-0.01]	1.8×10^{-3}	-0.01 [-0.02; 0.00]	1.1×10^{-1}	-0.07 [-0.29;0.15]	5.2×10^{-1}	-0.01 [-0.04; 0.02]	6.0×10^{-1}
ApoA1	0.03 [0.01; 0.05]	2.1×10^{-3}			0.07 [-0.17;0.31]	5.4×10^{-1}	0.02 [-0.01; 0.05]	1.7×10^{-1}
S-HDL-TG-%	-0.03 [-0.04;-0.01]	2.6×10^{-3}	0.01 [-0.00; 0.02]	1.0×10^{-1}	-0.12 [-0.35;0.11]	3.0×10^{-1}	-0.01 [-0.04; 0.02]	6.3×10^{-1}
GloI	-0.02 [-0.04;-0.01]	2.7×10^{-3}	-0.01 [-0.02;-0.00]	5.2×10^{-3}	0.18 [-0.18;0.54]	3.2×10^{-1}	-0.01 [-0.04; 0.03]	6.5×10^{-1}
S-HDL-TG	-0.02 [-0.04;-0.01]	3.6×10^{-3}	0.01 [-0.00; 0.02]	1.3×10^{-1}	-0.13 [-0.35;0.08]	2.3×10^{-1}	-0.01 [-0.04; 0.02]	3.9×10^{-1}
ApoB/ApoA1	-0.02 [-0.04;-0.01]	4.6×10^{-3}	-0.00 [-0.01; 0.01]	8.0×10^{-1}	-0.15 [-0.37;0.08]	2.0×10^{-1}	-0.01 [-0.04; 0.02]	5.6×10^{-1}
M-HDL-TG-%	-0.02 [-0.04;-0.01]	5.9×10^{-3}	0.01 [-0.00; 0.02]	2.2×10^{-1}	-0.13 [-0.35;0.09]	2.4×10^{-1}	-0.00 [-0.03; 0.03]	9.3×10^{-1}
M-HDL-CE	0.02 [0.01; 0.04]	7.1×10^{-3}	0.00 [-0.01; 0.01]	6.5×10^{-1}	0.12 [-0.13;0.36]	3.5×10^{-1}	0.02 [-0.01; 0.05]	1.5×10^{-1}
M-HDL-C	0.02 [0.01; 0.04]	7.7×10^{-3}	0.00 [-0.01; 0.01]	6.8×10^{-1}	0.12 [-0.13;0.36]	3.4×10^{-1}	0.02 [-0.01; 0.05]	1.6×10^{-1}
XL-HDL-TG-%	-0.02 [-0.04;-0.01]	7.9×10^{-3}	0.01 [-0.00; 0.02]	9.2×10^{-2}	-0.02 [-0.26;0.22]	8.9×10^{-1}	-0.01 [-0.04; 0.02]	4.0×10^{-1}
TG/PG	-0.02 [-0.04;-0.01]	9.8×10^{-3}	0.00 [-0.01; 0.01]	8.4×10^{-1}	-0.14 [-0.37;0.09]	2.3×10^{-1}	-0.02 [-0.05; 0.01]	2.1×10^{-1}
S-VLDL-P	-0.02 [-0.04;-0.00]	1.1×10^{-2}	-0.00 [-0.01; 0.01]	5.6×10^{-1}	-0.12 [-0.35;0.10]	2.9×10^{-1}	-0.01 [-0.04; 0.02]	4.0×10^{-1}
S-VLDL-TG	-0.02 [-0.04;-0.00]	1.1×10^{-2}	-0.00 [-0.01; 0.01]	6.2×10^{-1}	-0.12 [-0.34;0.11]	3.2×10^{-1}	-0.02 [-0.05; 0.01]	2.5×10^{-1}
S-VLDL-L	-0.02 [-0.04;-0.00]	1.2×10^{-2}	-0.01 [-0.02; 0.00]	2.1×10^{-1}	-0.14 [-0.36;0.09]	2.4×10^{-1}	-0.01 [-0.04; 0.02]	4.2×10^{-1}
S-VLDL-PL	-0.02 [-0.04;-0.00]	1.3×10^{-2}	-0.01 [-0.02; 0.00]	2.3×10^{-1}	-0.14 [-0.36;0.08]	2.2×10^{-1}	-0.01 [-0.04; 0.02]	5.3×10^{-1}
M-VLDL-PL	-0.02 [-0.04;-0.00]	1.4×10^{-2}	-0.00 [-0.01; 0.01]	7.4×10^{-1}	-0.10 [-0.33;0.12]	3.6×10^{-1}	-0.01 [-0.04; 0.02]	3.8×10^{-1}
XS-VLDL-TG	-0.02 [-0.04;-0.00]	1.4×10^{-2}	-0.00 [-0.01; 0.01]	9.7×10^{-1}	-0.13 [-0.37;0.10]	2.6×10^{-1}	-0.01 [-0.04; 0.02]	4.4×10^{-1}
M-HDL-FC	0.02 [0.00; 0.04]	1.4×10^{-2}			0.12 [-0.12;0.36]	3.3×10^{-1}	0.02 [-0.01; 0.05]	1.9×10^{-1}
M-VLDL-P	-0.02 [-0.04;-0.00]	1.5×10^{-2}	-0.00 [-0.01; 0.01]	6.0×10^{-1}	-0.10 [-0.32;0.13]	3.9×10^{-1}	-0.01 [-0.04; 0.01]	3.2×10^{-1}
S-VLDL-FC	-0.02 [-0.04;-0.00]	1.5×10^{-2}	-0.01 [-0.02; 0.00]	1.9×10^{-1}	-0.13 [-0.36;0.10]	2.6×10^{-1}	-0.01 [-0.04; 0.02]	3.8×10^{-1}
M-VLDL-TG	-0.02 [-0.04;-0.00]	1.7×10^{-2}	0.00 [-0.01; 0.01]	7.5×10^{-1}	-0.10 [-0.32;0.13]	4.0×10^{-1}	-0.02 [-0.05; 0.01]	2.0×10^{-1}
M-VLDL-L	-0.02 [-0.04;-0.00]	1.7×10^{-2}	-0.00 [-0.01; 0.01]	9.2×10^{-1}	-0.10 [-0.33;0.12]	3.7×10^{-1}	-0.01 [-0.04; 0.01]	3.4×10^{-1}
M-VLDL-FC	-0.02 [-0.04;-0.00]	1.9×10^{-2}	-0.00 [-0.01; 0.01]	8.1×10^{-1}	-0.10 [-0.32;0.13]	3.9×10^{-1}	-0.02 [-0.05; 0.01]	2.7×10^{-1}
M-HDL-L	0.02 [0.00; 0.04]	1.9×10^{-2}			0.10 [-0.15;0.35]	4.3×10^{-1}	0.03 [-0.00; 0.05]	6.4×10^{-2}
XS-VLDL-C-%	0.02 [0.00; 0.04]	2.0×10^{-2}	0.00 [-0.01; 0.01]	8.5×10^{-1}	0.03 [-0.18;0.24]	7.7×10^{-1}	0.01 [-0.02; 0.04]	4.9×10^{-1}
L-LDL-FC-%	0.02 [0.00; 0.04]	2.4×10^{-2}	-0.00 [-0.01; 0.01]	5.0×10^{-1}	0.15 [-0.08;0.39]	1.9×10^{-1}	0.00 [-0.03; 0.03]	7.9×10^{-1}
M-HDL-PL	0.02 [0.00; 0.04]	2.4×10^{-2}			0.09 [-0.16;0.34]	4.7×10^{-1}	0.02 [-0.01; 0.05]	1.9×10^{-1}
M-VLDL-C	-0.02 [-0.04;-0.00]	2.5×10^{-2}	-0.00 [-0.01; 0.01]	7.7×10^{-1}	-0.10 [-0.32;0.12]	3.8×10^{-1}	-0.01 [-0.04; 0.02]	5.4×10^{-1}
L-VLDL-PL	-0.02 [-0.03;-0.00]	2.6×10^{-2}	0.00 [-0.01; 0.01]	9.9×10^{-1}	-0.10 [-0.33;0.12]	3.7×10^{-1}	-0.00 [-0.03; 0.03]	9.4×10^{-1}
M-HDL-P	0.02 [0.00; 0.04]	2.6×10^{-2}			0.10 [-0.15;0.35]	4.3×10^{-1}	0.02 [-0.01; 0.05]	1.3×10^{-1}
L-VLDL-C	-0.02 [-0.03;-0.00]	2.7×10^{-2}	0.00 [-0.01; 0.01]	5.1×10^{-1}	-0.10 [-0.32;0.12]	3.7×10^{-1}	-0.00 [-0.03; 0.03]	8.9×10^{-1}
L-VLDL-TG	-0.02 [-0.03;-0.00]	3.3×10^{-2}	0.00 [-0.01; 0.01]	5.6×10^{-1}	-0.10 [-0.32;0.13]	4.0×10^{-1}	0.00 [-0.03; 0.03]	8.2×10^{-1}
L-VLDL-L	-0.02 [-0.03;-0.00]	3.3×10^{-2}	0.00 [-0.01; 0.01]	4.3×10^{-1}	-0.10 [-0.32;0.13]	3.9×10^{-1}	-0.00 [-0.03; 0.03]	9.9×10^{-1}
S-VLDL-C	-0.02 [-0.03;-0.00]	3.4×10^{-2}	-0.01 [-0.02;-0.00]	1.5×10^{-2}	-0.12 [-0.35;0.10]	2.9×10^{-1}	-0.01 [-0.04; 0.02]	5.8×10^{-1}
MUFA/FA	-0.02 [-0.03;-0.00]	4.0×10^{-2}	-0.01 [-0.02; 0.00]	2.7×10^{-1}	-0.15 [-0.38;0.08]	2.0×10^{-1}	-0.02 [-0.04; 0.01]	2.9×10^{-1}
XS-VLDL-CE-%	0.02 [0.00; 0.03]	4.2×10^{-2}	0.00 [-0.01; 0.01]	3.9×10^{-1}	0.03 [-0.17;0.24]	7.5×10^{-1}	0.02 [-0.01; 0.05]	1.7×10^{-1}
M-VLDL-CE	-0.02 [-0.03;-0.00]	4.4×10^{-2}	-0.00 [-0.01; 0.01]	7.7×10^{-1}	-0.10 [-0.31;0.12]	3.9×10^{-1}	-0.00 [-0.03; 0.02]	7.9×10^{-1}
VLDL-TG	-0.02 [-0.03;-0.00]	4.5×10^{-2}	0.00 [-0.01; 0.01]	7.6×10^{-1}	-0.13 [-0.35;0.09]	2.5×10^{-1}	-0.02 [-0.05; 0.01]	2.8×10^{-1}
Serum-TG	-0.02 [-0.03; 0.00]	5.5×10^{-2}	0.00 [-0.01; 0.01]	8.4×10^{-1}	-0.12 [-0.34;0.10]	2.9×10^{-1}	-0.01 [-0.04; 0.02]	4.5×10^{-1}
XS-VLDL-PL-%	0.02 [-0.00; 0.03]	5.7×10^{-2}	0.00 [-0.01; 0.01]	6.3×10^{-1}	0.05 [-0.17;0.28]	6.4×10^{-1}	0.00 [-0.02; 0.03]	7.6×10^{-1}
IDL-TG	-0.02 [-0.03; 0.00]	5.7×10^{-2}			-0.12 [-0.36;0.11]	3.1×10^{-1}	-0.02 [-0.04; 0.01]	2.8×10^{-1}
VLDL-C	-0.02 [-0.03; 0.00]	5.8×10^{-2}	-0.01 [-0.02; 0.00]	2.9×10^{-1}	-0.16 [-0.38;0.06]	1.5×10^{-1}	-0.00 [-0.03; 0.02]	8.1×10^{-1}
S-VLDL-CE	-0.02 [-0.03; 0.00]	7.6×10^{-2}	-0.01 [-0.02;-0.00]	1.1×10^{-2}	-0.10 [-0.32;0.12]	3.7×10^{-1}	-0.00 [-0.03; 0.02]	7.7×10^{-1}
L-VLDL-FC	-0.01 [-0.03; 0.00]	8.4×10^{-2}	0.00 [-0.01; 0.01]	8.3×10^{-1}	-0.10 [-0.32;0.13]	4.1×10^{-1}	-0.00 [-0.03; 0.03]	9.7×10^{-1}
Gp	-0.01 [-0.03; 0.00]	8.5×10^{-2}			-0.11 [-0.34;0.11]	3.3×10^{-1}	-0.01 [-0.04; 0.02]	4.6×10^{-1}
XS-VLDL-TG-%	-0.01 [-0.03; 0.00]	8.9×10^{-2}	0.00 [-0.01; 0.01]	9.3×10^{-1}	-0.02 [-0.24;0.21]	8.8×10^{-1}	-0.02 [-0.05; 0.01]	1.7×10^{-1}
M-LDL-TG	-0.01 [-0.03; 0.00]	9.3×10^{-2}			-0.13 [-0.35;0.09]	2.5×10^{-1}	-0.01 [-0.04; 0.02]	3.7×10^{-1}
MUFA	-0.01 [-0.03; 0.00]	1.2×10^{-1}	-0.00 [-0.01; 0.01]	4.8×10^{-1}	-0.15 [-0.38;0.08]	1.9×10^{-1}	-0.00 [-0.03; 0.02]	7.5×10^{-1}
IDL-TG-%	-0.01 [-0.03; 0.00]	1.3×10^{-1}			-0.05 [-0.29;0.19]	6.6×10^{-1}	-0.02 [-0.05; 0.01]	1.3×10^{-1}

continued on next page ...

XXL-VLDL-C-%	-0.01[-0.03: 0.00]	1.6×10^{-1}	-0.01[-0.02: 0.00]	6.5×10^{-2}	0.07[-0.19:0.34]	5.9×10^{-1}	-0.02[-0.05: 0.01]	1.1×10^{-1}
L-LDL-TG	-0.01[-0.03: 0.01]	1.7×10^{-1}	-0.00[-0.01: 0.01]	3.8×10^{-1}	-0.10[-0.32:0.13]	4.0×10^{-1}	-0.01[-0.04: 0.02]	5.3×10^{-1}
XS-VLDL-P	-0.01[-0.03: 0.01]	1.7×10^{-1}	-0.01[-0.02: 0.00]	1.9×10^{-1}	-0.12[-0.35:0.11]	3.0×10^{-1}	0.00[-0.03: 0.03]	9.9×10^{-1}
M-LDL-TG-%	-0.01[-0.03: 0.01]	1.9×10^{-1}			-0.05[-0.29:0.20]	7.2×10^{-1}	-0.01[-0.04: 0.02]	5.0×10^{-1}
LDL-TG	-0.01[-0.03: 0.01]	2.0×10^{-1}			-0.11[-0.33:0.11]	3.3×10^{-1}	-0.01[-0.04: 0.02]	4.7×10^{-1}
XS-VLDL-L	-0.01[-0.03: 0.01]	2.3×10^{-1}	-0.01[-0.02: 0.00]	2.0×10^{-1}	-0.13[-0.36:0.10]	2.8×10^{-1}	-0.00[-0.03: 0.03]	9.8×10^{-1}
XXL-VLDL-TG-%	0.01[-0.01: 0.03]	2.3×10^{-1}	0.01[0.00: 0.02]	3.2×10^{-2}	-0.01[-0.27:0.26]	9.5×10^{-1}	0.02[-0.00: 0.05]	8.7×10^{-2}
S-LDL-TG	-0.01[-0.03: 0.01]	2.5×10^{-1}			-0.13[-0.36:0.09]	2.5×10^{-1}	-0.01[-0.04: 0.02]	4.0×10^{-1}
L-LDL-TG-%	-0.01[-0.03: 0.01]	2.9×10^{-1}			-0.02[-0.26:0.22]	8.5×10^{-1}	-0.01[-0.04: 0.02]	4.4×10^{-1}
S-LDL-TG-%	-0.01[-0.03: 0.01]	3.0×10^{-1}	-0.01[-0.02: 0.00]	1.7×10^{-1}	-0.08[-0.33:0.17]	5.4×10^{-1}	-0.01[-0.04: 0.02]	4.0×10^{-1}
Remnant-C	-0.01[-0.03: 0.01]	3.1×10^{-1}	-0.01[-0.02: 0.00]	1.6×10^{-1}	-0.13[-0.36:0.10]	2.7×10^{-1}	0.00[-0.02: 0.03]	8.2×10^{-1}
S-LDL-C-%	0.01[-0.01: 0.02]	3.5×10^{-1}			-0.07[-0.30:0.17]	5.7×10^{-1}	-0.02[-0.04: 0.01]	2.6×10^{-1}
IDL-PL-%	0.01[-0.01: 0.02]	3.5×10^{-1}			0.16[-0.07:0.38]	1.7×10^{-1}	0.00[-0.03: 0.03]	9.2×10^{-1}
L-LDL-CE-%	0.01[-0.01: 0.03]	3.6×10^{-1}			-0.10[-0.32:0.11]	3.5×10^{-1}	0.00[-0.02: 0.03]	7.4×10^{-1}
XS-VLDL-FC	-0.01[-0.02: 0.01]	3.7×10^{-1}	-0.01[-0.02: 0.00]	6.1×10^{-2}	-0.11[-0.34:0.12]	3.6×10^{-1}	-0.01[-0.04: 0.02]	5.5×10^{-1}
Serum-C	0.01[-0.01: 0.03]	3.8×10^{-1}	-0.01[-0.02: 0.00]	9.0×10^{-2}	-0.04[-0.27:0.19]	7.2×10^{-1}	0.01[-0.02: 0.03]	6.2×10^{-1}
M-VLDL-FC-%	-0.01[-0.02: 0.01]	3.8×10^{-1}	-0.01[-0.02: 0.00]	7.9×10^{-2}	-0.05[-0.26:0.16]	6.4×10^{-1}	-0.02[-0.05: 0.01]	1.5×10^{-1}
S-LDL-CE-%	0.01[-0.01: 0.02]	4.2×10^{-1}			-0.07[-0.29:0.15]	5.1×10^{-1}	-0.01[-0.04: 0.01]	3.5×10^{-1}
Alb	0.01[-0.01: 0.02]	4.2×10^{-1}			0.35[0.12:0.58]	3.4×10^{-3}	0.01[-0.01: 0.04]	4.0×10^{-1}
L-LDL-C-%	0.01[-0.01: 0.02]	4.3×10^{-1}			-0.07[-0.30:0.15]	5.2×10^{-1}	-0.00[-0.03: 0.02]	8.5×10^{-1}
EstC	0.01[-0.01: 0.02]	4.4×10^{-1}	-0.01[-0.02: 0.00]	7.4×10^{-2}	-0.03[-0.26:0.21]	8.3×10^{-1}	0.01[-0.02: 0.04]	5.2×10^{-1}
M-LDL-PL	-0.01[-0.02: 0.01]	4.5×10^{-1}	-0.01[-0.02: 0.00]	3.0×10^{-1}	-0.08[-0.31:0.15]	4.9×10^{-1}	0.00[-0.02: 0.03]	8.8×10^{-1}
XS-VLDL-PL	-0.01[-0.02: 0.01]	4.7×10^{-1}	-0.01[-0.02: 0.00]	2.2×10^{-1}	-0.10[-0.33:0.13]	3.9×10^{-1}	0.00[-0.02: 0.03]	8.4×10^{-1}
IDL-C-%	0.01[-0.01: 0.02]	4.7×10^{-1}			-0.03[-0.26:0.21]	8.2×10^{-1}	0.02[-0.01: 0.05]	1.8×10^{-1}
HDL-TG	-0.01[-0.02: 0.01]	4.8×10^{-1}	0.00[-0.01: 0.01]	5.1×10^{-1}	-0.09[-0.32:0.14]	4.6×10^{-1}	0.00[-0.02: 0.03]	7.7×10^{-1}
XS-VLDL-C	-0.01[-0.02: 0.01]	5.2×10^{-1}			-0.10[-0.33:0.12]	3.7×10^{-1}	0.00[-0.02: 0.03]	7.5×10^{-1}
M-LDL-C-%	0.00[-0.01: 0.02]	5.6×10^{-1}			-0.05[-0.27:0.17]	6.7×10^{-1}	-0.02[-0.04: 0.01]	2.4×10^{-1}
IDL-CE-%	0.00[-0.01: 0.02]	6.0×10^{-1}			-0.04[-0.26:0.18]	7.3×10^{-1}	0.03[0.00: 0.05]	4.9×10^{-2}
M-LDL-L	-0.00[-0.02: 0.01]	6.6×10^{-1}	-0.00[-0.01: 0.01]	3.4×10^{-1}	-0.07[-0.30:0.15]	5.3×10^{-1}	-0.00[-0.03: 0.02]	7.7×10^{-1}
IDL-FC	0.00[-0.01: 0.02]	6.8×10^{-1}	-0.01[-0.02: 0.00]	2.5×10^{-1}	-0.01[-0.25:0.22]	9.1×10^{-1}	0.00[-0.02: 0.03]	7.6×10^{-1}
XXL-VLDL-FC-%	0.00[-0.01: 0.02]	7.0×10^{-1}	0.00[-0.01: 0.01]	4.6×10^{-1}	-0.19[-0.46:0.08]	1.8×10^{-1}	-0.01[-0.04: 0.01]	3.3×10^{-1}
M-LDL-FC	-0.00[-0.02: 0.01]	7.1×10^{-1}	-0.01[-0.02: 0.00]	3.0×10^{-1}	-0.06[-0.29:0.17]	6.2×10^{-1}	-0.00[-0.03: 0.02]	8.9×10^{-1}
S-LDL-PL	0.00[-0.01: 0.02]	7.4×10^{-1}	-0.00[-0.01: 0.01]	5.1×10^{-1}	-0.07[-0.30:0.16]	5.6×10^{-1}	0.00[-0.02: 0.03]	8.3×10^{-1}
L-LDL-FC	0.00[-0.01: 0.02]	7.6×10^{-1}	-0.00[-0.01: 0.00]	3.3×10^{-1}	-0.01[-0.25:0.22]	9.2×10^{-1}	0.00[-0.02: 0.03]	8.6×10^{-1}
L-LDL-P	-0.00[-0.02: 0.02]	8.2×10^{-1}	-0.00[-0.01: 0.01]	5.0×10^{-1}	-0.04[-0.27:0.19]	7.4×10^{-1}	0.00[-0.02: 0.03]	7.6×10^{-1}
S-LDL-FC	0.00[-0.02: 0.02]	8.3×10^{-1}	-0.01[-0.02: 0.00]	1.9×10^{-1}	-0.05[-0.28:0.18]	6.7×10^{-1}	-0.00[-0.03: 0.02]	7.2×10^{-1}
M-LDL-C	-0.00[-0.02: 0.02]	8.4×10^{-1}	-0.01[-0.02: 0.00]	2.3×10^{-1}	-0.05[-0.28:0.18]	6.7×10^{-1}	-0.00[-0.03: 0.02]	8.2×10^{-1}
IDL-C	0.00[-0.02: 0.02]	8.5×10^{-1}	-0.01[-0.02: 0.00]	1.1×10^{-1}	-0.03[-0.27:0.20]	7.8×10^{-1}	0.01[-0.02: 0.03]	6.2×10^{-1}
M-LDL-CE	-0.00[-0.02: 0.02]	8.7×10^{-1}	-0.01[-0.02: 0.00]	2.2×10^{-1}	-0.04[-0.27:0.18]	7.1×10^{-1}	-0.00[-0.03: 0.02]	8.4×10^{-1}
LDL-C	0.00[-0.02: 0.02]	9.0×10^{-1}	-0.01[-0.02: 0.00]	2.5×10^{-1}	-0.09[-0.32:0.14]	4.5×10^{-1}	-0.00[-0.03: 0.02]	8.5×10^{-1}
M-LDL-CE-%	-0.00[-0.02: 0.02]	9.1×10^{-1}			-0.04[-0.26:0.17]	7.0×10^{-1}	-0.01[-0.04: 0.02]	4.8×10^{-1}
L-LDL-CE	-0.00[-0.02: 0.02]	9.3×10^{-1}	-0.01[-0.02: 0.00]	2.4×10^{-1}	-0.05[-0.28:0.18]	6.9×10^{-1}	0.00[-0.02: 0.03]	8.6×10^{-1}
IDL-P	-0.00[-0.02: 0.02]	9.3×10^{-1}	-0.01[-0.02: 0.00]	2.6×10^{-1}	-0.07[-0.30:0.15]	5.3×10^{-1}	0.00[-0.02: 0.03]	8.1×10^{-1}
S-LDL-L	0.00[-0.02: 0.02]	9.5×10^{-1}	-0.01[-0.02: 0.00]	2.6×10^{-1}	-0.06[-0.29:0.16]	5.9×10^{-1}	-0.00[-0.03: 0.02]	8.1×10^{-1}
IDL-L	0.00[-0.02: 0.02]	9.6×10^{-1}	-0.01[-0.02: 0.00]	2.3×10^{-1}	-0.07[-0.30:0.16]	5.6×10^{-1}	0.00[-0.02: 0.03]	7.9×10^{-1}
S-LDL-CE	-0.00[-0.02: 0.02]	9.6×10^{-1}	-0.01[-0.02: 0.00]	1.3×10^{-1}	-0.03[-0.26:0.19]	7.7×10^{-1}	-0.00[-0.03: 0.02]	8.9×10^{-1}
L-LDL-L	-0.00[-0.02: 0.02]	9.6×10^{-1}	-0.00[-0.01: 0.01]	3.4×10^{-1}	-0.05[-0.28:0.18]	6.6×10^{-1}	0.00[-0.03: 0.03]	9.5×10^{-1}
L-LDL-C	0.00[-0.02: 0.02]	9.6×10^{-1}	-0.01[-0.02: 0.00]	2.5×10^{-1}	-0.04[-0.27:0.19]	7.1×10^{-1}	0.00[-0.02: 0.03]	9.0×10^{-1}
IDL-CE	0.00[-0.02: 0.02]	9.7×10^{-1}	-0.01[-0.02: 0.00]	7.9×10^{-2}	-0.04[-0.27:0.20]	7.6×10^{-1}	0.01[-0.02: 0.04]	5.1×10^{-1}
IDL-PL	-0.00[-0.02: 0.02]	9.8×10^{-1}	-0.01[-0.02: 0.00]	3.0×10^{-1}	-0.03[-0.26:0.20]	7.9×10^{-1}	0.00[-0.02: 0.03]	7.3×10^{-1}
S-LDL-C	0.00[-0.02: 0.02]	9.8×10^{-1}	-0.01[-0.02: 0.00]	1.4×10^{-1}	-0.04[-0.27:0.19]	7.3×10^{-1}	-0.00[-0.03: 0.02]	8.3×10^{-1}
L-LDL-PL	-0.00[-0.02: 0.02]	9.8×10^{-1}	-0.01[-0.02: 0.00]	1.9×10^{-1}	-0.03[-0.27:0.20]	8.0×10^{-1}	0.01[-0.02: 0.03]	6.6×10^{-1}
M-HDL-FC-%					0.15[-0.09:0.40]	2.3×10^{-1}	0.01[-0.02: 0.04]	3.5×10^{-1}
M-HDL-C-%					0.14[-0.09:0.36]	2.4×10^{-1}	0.01[-0.02: 0.04]	4.6×10^{-1}
XL-HDL-PL-%					0.14[-0.10:0.39]	2.6×10^{-1}	0.04[0.01: 0.07]	3.7×10^{-3}
Val					0.12[-0.11:0.36]	3.1×10^{-1}	0.00[-0.03: 0.03]	9.1×10^{-1}
Ala					0.12[-0.11:0.35]	3.2×10^{-1}	0.00[-0.03: 0.03]	9.1×10^{-1}
L-VLDL-CE					-0.11[-0.33:0.11]	3.4×10^{-1}	-0.00[-0.03: 0.03]	8.8×10^{-1}
L-HDL-FC-%					0.10[-0.11:0.30]	3.6×10^{-1}	0.02[-0.01: 0.05]	1.9×10^{-1}
XL-HDL-C-%					-0.11[-0.36:0.14]	3.8×10^{-1}	-0.04[-0.07:0.01]	1.1×10^{-2}
XL-VLDL-CE					-0.10[-0.32:0.12]	3.8×10^{-1}	-0.01[-0.03: 0.02]	7.2×10^{-1}
XL-VLDL-C					-0.10[-0.32:0.13]	3.9×10^{-1}	-0.00[-0.03: 0.02]	7.8×10^{-1}
Phe					-0.11[-0.36:0.14]	4.0×10^{-1}	-0.02[-0.05: 0.01]	2.4×10^{-1}
M-HDL-PL-%					-0.10[-0.33:0.14]	4.2×10^{-1}	-0.03[-0.05: 0.00]	6.8×10^{-2}
L-HDL-CE-%					0.06[-0.15:0.27]	5.5×10^{-1}	0.01[-0.02: 0.04]	4.6×10^{-1}
XL-HDL-CE-%					-0.07[-0.32:0.18]	6.0×10^{-1}	-0.05[-0.07:0.02]	2.1×10^{-3}
Leu					0.05[-0.19:0.28]	7.0×10^{-1}	-0.01[-0.04: 0.03]	7.4×10^{-1}
SFA/FA					0.04[-0.17:0.25]	7.1×10^{-1}	0.01[-0.02: 0.03]	5.9×10^{-1}
XXL-VLDL-C					-0.04[-0.27:0.19]	7.5×10^{-1}	-0.00[-0.03: 0.03]	9.8×10^{-1}
Pyr					-0.03[-0.25:0.19]	7.9×10^{-1}	0.01[-0.01: 0.04]	3.3×10^{-1}
L-HDL-TG-%					-0.02[-0.24:0.21]	8.8×10^{-1}	-0.01[-0.04: 0.01]	3.7×10^{-1}
XXL-VLDL-CE					-0.01[-0.24:0.21]	9.2×10^{-1}	0.00[-0.03: 0.03]	9.1×10^{-1}
Ile					0.01[-0.21:0.24]	9.2×10^{-1}	-0.00[-0.04: 0.03]	8.9×10^{-1}
Gly					0.01[-0.24:0.25]	9.6×10^{-1}	-0.04[-0.07:0.01]	2.1×10^{-2}
Lac					-0.00[-0.21:0.21]	9.9×10^{-1}	0.01[-0.02: 0.04]	6.5×10^{-1}

Table C.3 Associations of NMR metabolite with renal function. I regressed 227 metabolic traits against the estimated glomerular filtration rate (eGFR) in three cohorts of type 2 diabetics and three cohorts of non-diabetics. Fixed effect inverse meta-analysis was used to combine results for both groups.

ID	Metabolite	Diabetics										Non-Diabetics									
		GDM		TwinsUK		KORA		Meta		TwinsUK		KORA		YoungFinns		Meta					
		β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p				
XXL-VLDL-P	Concentration of chylomicrons and extremely large VLDL particles	1.74[-0.33: 3.82]	1.0×10 ⁻¹	1.57[-1.27: 4.40]	2.8×10 ⁻¹	-1.53[-4.14: 1.09]	2.5×10 ⁻¹	0.75[-0.66: 2.16]	3.0×10 ⁻¹	-0.84[-1.59:-0.09]	2.9×10 ⁻²	-0.49[-1.12: 0.14]	1.3×10 ⁻¹	-0.02[-0.59: 0.56]	9.5×10 ⁻¹	-0.38[-0.75:-0.01]	4.5×10 ⁻²				
XXL-VLDL-L	Total lipids in chylomicrons and extremely large VLDL	1.69[-0.39: 3.76]	1.1×10 ⁻¹	1.55[-1.28: 4.39]	2.8×10 ⁻¹	-1.59[-4.20: 1.03]	2.4×10 ⁻¹	0.70[-0.71: 2.11]	3.3×10 ⁻¹	-0.85[-1.60:-0.10]	2.7×10 ⁻²	-0.50[-1.13: 0.13]	1.2×10 ⁻¹	-0.05[-0.62: 0.53]	8.7×10 ⁻¹	-0.40[-0.77:-0.03]	3.5×10 ⁻²				
XXL-VLDL-PL	Phospholipids in chylomicrons and extremely large VLDL	1.50[-0.57: 3.58]	1.6×10 ⁻¹	1.46[-1.38: 4.31]	3.1×10 ⁻¹	-1.69[-4.29: 0.92]	2.1×10 ⁻¹	0.56[-0.85: 1.97]	4.4×10 ⁻¹	-0.49[-1.24: 0.26]	2.0×10 ⁻¹	-0.54[-1.17: 0.09]	9.3×10 ⁻²	-0.06[-0.63: 0.51]	8.5×10 ⁻¹	-0.33[-0.69: 0.04]	8.3×10 ⁻²				
XXL-VLDL-C	Total cholesterol in chylomicrons and extremely large VLDL	0.10[-1.98: 2.17]	9.3×10 ⁻¹	1.21[-1.59: 4.00]	4.0×10 ⁻¹	-1.93[-4.54: 0.67]	1.5×10 ⁻¹	-0.21[-1.62: 1.19]	7.7×10 ⁻¹	-1.09[-1.84:-0.35]	4.1×10 ⁻³	-0.74[-1.36:-0.11]	2.1×10 ⁻²	-0.32[-0.89: 0.25]	2.7×10 ⁻¹	-0.65[-1.02:-0.28]	5.2×10 ⁻⁴				
XXL-VLDL-CE	Cholesterol esters in chylomicrons and extremely large VLDL	-0.41[-2.48: 1.67]	7.0×10 ⁻¹	1.28[-1.47: 4.02]	3.6×10 ⁻¹	-1.83[-4.43: 0.77]	1.7×10 ⁻¹	-0.38[-1.78: 1.02]	5.9×10 ⁻¹	-1.05[-1.80:-0.31]	5.5×10 ⁻³	-0.67[-1.29:-0.05]	3.5×10 ⁻²	-0.42[-0.99: 0.14]	1.4×10 ⁻¹	-0.66[-1.02:-0.29]	3.9×10 ⁻⁴				
XXL-VLDL-FC	Free cholesterol in chylomicrons and extremely large VLDL	0.90[-1.18: 2.98]	4.0×10 ⁻¹	1.03[-1.82: 3.87]	4.8×10 ⁻¹	-2.00[-4.60: 0.61]	1.3×10 ⁻¹	0.08[-1.33: 1.49]	9.1×10 ⁻¹	-0.63[-1.38: 0.12]	1.0×10 ⁻¹	-0.59[-1.22: 0.04]	6.9×10 ⁻²	-0.10[-0.67: 0.48]	7.4×10 ⁻¹	-0.39[-0.76:-0.02]	3.7×10 ⁻²				
XXL-VLDL-TG	Triglycerides in chylomicrons and extremely large VLDL	2.00[-0.07: 4.08]	5.9×10 ⁻²	1.62[-1.22: 4.46]	2.6×10 ⁻¹	-1.45[-4.07: 1.17]	2.8×10 ⁻¹	0.90[-0.51: 2.32]	2.1×10 ⁻¹	-0.77[-1.52:-0.02]	4.5×10 ⁻²	-0.46[-1.09: 0.17]	1.5×10 ⁻¹	0.03[-0.55: 0.60]	9.3×10 ⁻¹	-0.33[-0.70: 0.04]	7.7×10 ⁻²				
XL-VLDL-P	Concentration of very large VLDL particles	0.59[-1.49: 2.67]	5.8×10 ⁻¹	1.37[-1.49: 4.23]	3.5×10 ⁻¹	-2.01[-4.61: 0.59]	1.3×10 ⁻¹	0.01[-1.40: 1.43]	9.8×10 ⁻¹	-0.75[-1.51: 0.01]	5.2×10 ⁻²	-0.68[-1.31:-0.04]	3.7×10 ⁻²	-0.20[-0.78: 0.38]	4.9×10 ⁻¹	-0.50[-0.87:-0.13]	8.7×10 ⁻³				
XL-VLDL-L	Total lipids in very large VLDL	0.68[-1.40: 2.76]	5.2×10 ⁻¹	1.38[-1.48: 4.24]	3.4×10 ⁻¹	-2.07[-4.67: 0.53]	1.2×10 ⁻¹	0.04[-1.37: 1.45]	9.6×10 ⁻¹	-0.77[-1.52;-0.01]	4.7×10 ⁻²	-0.72[-1.36;-0.08]	2.7×10 ⁻²	-0.24[-0.83: 0.34]	4.2×10 ⁻¹	-0.54[-0.91:-0.16]	5.1×10 ⁻³				
XL-VLDL-PL	Phospholipids in very large VLDL	0.88[-1.19: 2.96]	4.0×10 ⁻¹	1.27[-1.57: 4.11]	3.8×10 ⁻¹	-1.93[-4.53: 0.67]	1.5×10 ⁻¹	0.15[-1.26: 1.56]	8.3×10 ⁻¹	-0.72[-1.47: 0.03]	6.2×10 ⁻²	-0.67[-1.30;-0.03]	3.9×10 ⁻²	-0.16[-0.74: 0.42]	6.0×10 ⁻¹	-0.47[-0.84:-0.10]	1.3×10 ⁻²				
XL-VLDL-C	Total cholesterol in very large VLDL	0.22[-1.85: 2.30]	8.3×10 ⁻¹	0.96[-1.86: 3.78]	5.0×10 ⁻¹	-2.13[-4.73: 0.47]	1.1×10 ⁻¹	-0.28[-1.69: 1.12]	6.9×10 ⁻¹	-1.23[-1.98:-0.48]	1.3×10 ⁻³	-0.81[-1.44:-0.17]	1.3×10 ⁻²	-0.33[-0.91: 0.24]	2.6×10 ⁻¹	-0.71[-1.08;-0.34]	1.5×10 ⁻⁴				
XL-VLDL-CE	Cholesterol esters in very large VLDL	-0.36[-2.43: 1.71]	7.4×10 ⁻¹	0.91[-1.90: 3.73]	5.2×10 ⁻¹	-2.23[-4.83: 0.37]	9.5×10 ⁻²	-0.59[-1.99: 0.82]	4.1×10 ⁻¹	-1.29[-2.04;-0.54]	7.3×10 ⁻⁴	-0.86[-1.49;-0.23]	7.6×10 ⁻³	-0.36[-0.93: 0.21]	2.2×10 ⁻¹	-0.76[-1.13;-0.39]	5.8×10 ⁻⁵				
XL-VLDL-FC	Free cholesterol in very large VLDL	0.89[-1.18: 2.97]	4.0×10 ⁻¹	0.98[-1.85: 3.81]	5.0×10 ⁻¹	-1.98[-4.59: 0.62]	1.4×10 ⁻¹	0.07[-1.33: 1.48]	9.2×10 ⁻¹	-0.73[-1.48: 0.01]	5.4×10 ⁻²	-0.69[-1.32;-0.06]	3.3×10 ⁻²	-0.25[-0.83: 0.33]	3.9×10 ⁻¹	-0.52[-0.89;-0.15]	6.0×10 ⁻³				
XL-VLDL-TG	Triglycerides in very large VLDL	0.69[-1.39: 2.77]	5.1×10 ⁻¹	1.49[-1.38: 4.36]	3.1×10 ⁻¹	-2.03[-4.63: 0.57]	1.3×10 ⁻¹	0.08[-1.33: 1.49]	9.1×10 ⁻¹	-0.73[-1.49: 0.03]	6.0×10 ⁻²	-0.72[-1.36;-0.08]	2.7×10 ⁻²	-0.19[-0.78: 0.39]	5.2×10 ⁻¹	-0.51[-0.88;-0.13]	8.3×10 ⁻³				
L-VLDL-P	Concentration of large VLDL particles	-0.25[-2.32: 1.83]	8.2×10 ⁻¹	1.94[-0.98: 4.85]	1.9×10 ⁻¹	-2.16[-4.76: 0.44]	1.0×10 ⁻¹	-0.30[-1.72: 1.12]	6.8×10 ⁻¹	-0.75[-1.52: 0.02]	5.6×10 ⁻²	-0.75[-1.39;-0.11]	2.1×10 ⁻²	-0.34[-0.92: 0.24]	2.5×10 ⁻¹	-0.58[-0.96;-0.20]	2.5×10 ⁻³				
L-VLDL-L	Total lipids in large VLDL	0.06[-2.02: 2.14]	9.5×10 ⁻¹	2.08[-0.84: 5.00]	1.6×10 ⁻¹	-2.46[-5.04: 0.11]	9.3×10 ⁻²	-0.22[-1.64: 1.19]	7.6×10 ⁻¹	-1.03[-1.80;-0.25]	9.2×10 ⁻³	-0.94[-1.58;-0.29]	4.5×10 ⁻³	-0.28[-0.88: 0.31]	3.5×10 ⁻¹	-0.69[-1.07;-0.31]	3.7×10 ⁻⁴				
L-VLDL-PL	Phospholipids in large VLDL	-0.14[-2.22: 1.94]	8.9×10 ⁻¹	1.99[-0.91: 4.89]	1.8×10 ⁻¹	-2.21[-4.81: 0.39]	9.7×10 ⁻²	-0.25[-1.67: 1.17]	7.3×10 ⁻¹	-1.10[-1.87;-0.33]	4.9×10 ⁻³	-0.85[-1.49;-0.20]	9.9×10 ⁻³	-0.40[-0.99: 0.18]	1.8×10 ⁻¹	-0.72[-1.10;-0.35]	1.7×10 ⁻⁴				
L-VLDL-C	Total cholesterol in large VLDL	-0.48[-2.55: 1.60]	6.5×10 ⁻¹	1.80[-1.07: 4.68]	2.2×10 ⁻¹	-2.42[-5.01: 0.18]	7.0×10 ⁻²	-0.50[-1.92: 0.91]	4.8×10 ⁻¹	-1.12[-1.89;-0.36]	4.0×10 ⁻³	-0.94[-1.58;-0.30]	4.0×10 ⁻³	-0.42[-1.00: 0.17]	1.6×10 ⁻¹	-0.77[-1.14;-0.39]	6.1×10 ⁻⁵				
L-VLDL-CE	Cholesterol esters in large VLDL	-0.92[-3.00: 1.15]	3.8×10 ⁻¹	1.87[-1.00: 4.73]	2.0×10 ⁻¹	-2.44[-5.04: 0.16]	6.7×10 ⁻²	-0.70[-2.11: 0.71]	3.3×10 ⁻¹	-1.08[-1.85;-0.31]	5.7×10 ⁻³	-0.97[-1.60;-0.33]	2.8×10 ⁻³	-0.40[-0.98: 0.18]	1.8×10 ⁻¹	-0.76[-1.13;-0.38]	7.1×10 ⁻⁵				
L-VLDL-FC	Free cholesterol in large VLDL	-0.08[-2.16: 1.99]	9.4×10 ⁻¹	1.62[-1.26: 4.51]	2.7×10 ⁻¹	-2.31[-4.91: 0.28]	8.2×10 ⁻²	-0.34[-1.75: 1.08]	6.4×10 ⁻¹	-0.78[-1.54;-0.03]	4.3×10 ⁻²	-0.83[-1.47;-0.19]	1.1×10 ⁻²	-0.41[-0.99: 0.17]	1.7×10 ⁻¹	-0.64[-1.02;-0.27]	7.2×10 ⁻⁴				
L-VLDL-TG	Triglycerides in large VLDL	0.07[-2.01: 2.15]	9.5×10 ⁻¹	2.05[-0.88: 4.99]	1.7×10 ⁻¹	-2.19[-4.79: 0.41]	1.0×10 ⁻¹	-0.14[-1.56: 1.28]	8.5×10 ⁻¹	-1.02[-1.80;-0.25]	9.5×10 ⁻³	-0.88[-1.53;-0.24]	7.6×10 ⁻³	-0.25[-0.84: 0.35]	4.2×10 ⁻¹	-0.65[-1.03;-0.27]	7.6×10 ⁻⁴				
M-VLDL-P	Concentration of medium VLDL particles	-0.96[-3.03: 1.11]	3.6×10 ⁻¹	2.23[-0.69: 5.16]	1.3×10 ⁻¹	-2.69[-5.26;-0.11]	4.2×10 ⁻²	-0.74[-2.15: 0.68]	3.1×10 ⁻¹	-1.15[-1.92;-0.38]	3.3×10 ⁻³	-1.25[-1.90;-0.60]	1.7×10 ⁻⁴	-0.52[-1.11: 0.07]	8.2×10 ⁻²	-0.92[-1.30;-0.54]	1.8×10 ⁻⁶				
M-VLDL-L	Total lipids in medium VLDL	-0.50[-2.57: 1.57]	6.3×10 ⁻¹	2.49[-0.44: 5.42]	9.6×10 ⁻²	-2.68[-5.26;-0.11]	4.3×10 ⁻²	-0.47[-1.88: 0.95]	5.2×10 ⁻¹	-1.05[-1.82;-0.28]	7.3×10 ⁻³	-1.25[-1.90;-0.60]	1.8×10 ⁻⁴	-0.55[-1.16: 0.05]	7.2×10 ⁻²	-0.92[-1.30;-0.54]	2.6×10 ⁻⁶				
M-VLDL-PL	Phospholipids in medium VLDL	-0.75[-2.82: 1.32]	4.8×10 ⁻¹	2.34[-0.56: 5.23]	1.1×10 ⁻¹	-2.72[-5.28;-0.15]	3.9×10 ⁻²	-0.62[-2.02: 0.79]	3.9×10 ⁻¹	-1.14[-1.90;-0.37]	3.6×10 ⁻³	-1.25[-1.90;-0.60]	1.6×10 ⁻⁴	-0.55[-1.14: 0.04]	7.0×10 ⁻²	-0.93[-1.31;-0.55]	1.4×10 ⁻⁶				
M-VLDL-C	Total cholesterol in medium VLDL	-0.77[-2.84: 1.30]	4.6×10 ⁻¹	2.49[-0.32: 5.29]	8.2×10 ⁻²	-2.56[-5.12;-0.00]	5.1×10 ⁻²	-0.50[-1.89: 0.90]	4.8×10 ⁻¹	-0.98[-1.73;-0.23]	1.1×10 ⁻²	-1.13[-1.76;-0.49]	5.3×10 ⁻⁴	-0.60[-1.18;-0.01]	4.5×10 ⁻²	-0.87[-1.25;-0.50]	4.4×10 ⁻⁶				
M-VLDL-CE	Cholesterol esters in medium VLDL	-0.55[-2.62: 1.52]	6.0×10 ⁻¹	2.46[-0.28: 5.19]	7.8×10 ⁻²	-2.33[-4.88: 0.22]	7.5×10 ⁻²	-0.30[-1.69: 1.08]	6.7×10 ⁻¹	-0.82[-1.57;-0.08]	2.9×10 ⁻²	-0.92[-1.55;-0.30]	3.8×10 ⁻³	-0.52[-1.09: 0.06]	7.7×10 ⁻²	-0.73[-1.10;-0.37]	9.0×10 ⁻⁵				
M-VLDL-FC	Free cholesterol in medium VLDL	-1.07[-3.14: 1.00]	3.1×10 ⁻¹	2.22[-0.67: 5.11]	1.3×10 ⁻¹	-2.78[-5.35;-0.21]	3.5×10 ⁻²	-0.80[-2.21: 0.60]	2.6×10 ⁻¹	-1.15[-1.91;-0.38]	3.2×10 ⁻³	-1.34[-1.99;-0.69]	5.2×10 ⁻⁵	-0.55[-1.14: 0.04]	6.6×10 ⁻²	-0.97[-1.34;-0.59]	5.3×10 ⁻⁷				
M-VLDL-TG	Triglycerides in medium VLDL	-0.66[-2.73: 1.41]	5.3×10 ⁻¹	2.23[-0.76: 5.22]	1.4×10 ⁻¹	-2.76[-5.34;-0.18]	3.8×10 ⁻²	-0.64[-2.06: 0.78]	3.8×10 ⁻¹	-1.11[-1.89;-0.34]	5.0×10 ⁻³	-1.29[-1.94;-0.63]	1.2×10 ⁻⁴	-0.42[-1.02: 0.19]	1.8×10 ⁻¹	-0.89[-1.27;-0.50]	6.3×10 ⁻⁶				
S-VLDL-P	Concentration of small VLDL particles	-1.92[-3.97: 0.14]	6.8×10 ⁻²	2.43[-0.41: 5.28]	9.4×10 ⁻²	-3.00[-5.58;-0.42]	2.4×10 ⁻²	-1.18[-2.58: 0.22]	9.8×10 ⁻²	-1.16[-1.92;-0.39]	3.0×10 ⁻³	-1.54[-2.18;-0.89]	3.2×10 ⁻⁶	-0.98[-1.57;-0.38]	1.3×10 ⁻³	-1.22[-1.59;-0.84]	3.5×10 ⁻¹⁰				
S-VLDL-L	Total lipids in small VLDL	-1.44[-3.50: 0.62]	1.7×10 ⁻¹	2.43[-0.40: 5.27]	9.2×10 ⁻²	-2.98[-5.56;-0.40]	2.5×10 ⁻²	-0.95[-2.35: 0.45]	1.8×10 ⁻¹	-1.09[-1.85;-0.33]	5.1×10 ⁻³	-1.53[-2.17;-0.89]	3.3×10 ⁻⁶	-1.14[-1.74;-0.53]	2.2×10 ⁻⁴	-1.26[-1.64;-0.88]	8.0×10 ⁻¹¹				
S-VLDL-PL	Phospholipids in small VLDL	-1.78[-3.84: 0.28]	9.1×10 ⁻²	2.39[-0.45: 5.22]	9.9×10 ⁻²	-3.00[-5.58;-0.41]	2.4×10 ⁻														

L-LDL-PL	Phospholipids in large LDL	6.48[4.47:	8.48]	4.4×10 ⁻¹⁰	2.53[-0.13:	5.19]	6.2×10 ⁻²	0.65[-1.95:	3.25]	6.3×10 ⁻¹	3.84[2.47:	5.20]	3.4×10 ⁻⁸	0.08[-0.66:	0.82]	8.4×10 ⁻¹	0.04[-0.56:	0.64]	9.0×10 ⁻¹	-0.30[-0.84:	0.23]	2.7×10 ⁻¹	-0.10[-0.45:	0.25]	5.8×10 ⁻¹		
L-LDL-C	Total cholesterol in large LDL	6.33[4.33:	8.33]	9.9×10 ⁻¹⁰	2.45[-0.20:	5.10]	7.0×10 ⁻²	0.69[-1.91:	3.28]	6.1×10 ⁻¹	3.76[2.40:	5.12]	6.1×10 ⁻⁸	0.07[-0.67:	0.81]	8.5×10 ⁻¹	0.07[-0.53:	0.67]	8.3×10 ⁻¹	-0.28[-0.81:	0.26]	3.1×10 ⁻¹	-0.08[-0.43:	0.27]	6.5×10 ⁻¹		
L-LDL-CE	Cholesterol esters in large LDL	6.15[4.14:	8.15]	3.0×10 ⁻⁹	2.47[-0.18:	5.12]	6.8×10 ⁻²	0.51[-2.09:	3.10]	7.0×10 ⁻¹	3.63[2.26:	4.99]	1.8×10 ⁻⁷	0.03[-0.70:	0.77]	9.3×10 ⁻¹	0.02[-0.58:	0.62]	9.6×10 ⁻¹	-0.29[-0.82:	0.25]	3.0×10 ⁻¹	-0.11[-0.46:	0.24]	5.4×10 ⁻¹		
L-LDL-FC	Free cholesterol in large LDL	6.34[4.34:	8.34]	9.2×10 ⁻¹⁰	2.40[-0.29:	5.08]	8.0×10 ⁻²	1.01[-1.60:	3.62]	4.5×10 ⁻¹	3.86[2.49:	5.22]	3.1×10 ⁻⁸	0.18[-0.56:	0.93]	6.3×10 ⁻¹	0.19[-0.41:	0.80]	5.3×10 ⁻¹	-0.26[-0.78:	0.27]	3.4×10 ⁻¹	-0.01[-0.36:	0.34]	9.6×10 ⁻¹		
L-LDL-TG	Triglycerides in large LDL	-1.96[-4.03:	0.12]	6.5×10 ⁻²	2.61[-0.02:	5.24]	5.1×10 ⁻²	-1.87[-4.49:	0.75]	1.6×10 ⁻¹	-0.67[-2.05:	0.72]	3.5×10 ⁻¹	-0.67[-1.41:	0.07]	7.4×10 ⁻²	-1.31[-1.92:-0.71]	2.3×10 ⁻⁵	-0.58[-1.12:-0.05]	3.3×10 ⁻²	-0.85[-1.20:-0.50]	2.3×10 ⁻⁶						
M-LDL-P	Concentration of medium LDL particles	4.92[2.89:	6.94]	2.4×10 ⁻⁶	2.56[-0.08:	5.21]	5.7×10 ⁻²	-0.17[-2.78:	2.44]	9.0×10 ⁻¹	2.89[1.52:	4.26]	3.5×10 ⁻⁵	-0.08[-0.82:	0.65]	8.2×10 ⁻¹	-0.10[-0.70:	0.50]	7.4×10 ⁻¹	-0.35[-0.90:	0.19]	2.0×10 ⁻¹	-0.20[-0.56:	0.15]	2.6×10 ⁻¹		
M-LDL-L	Total lipids in medium LDL	5.55[3.54:	7.56]	9.3×10 ⁻⁸	2.51[-0.12:	5.15]	6.2×10 ⁻²	0.16[-2.45:	2.76]	9.1×10 ⁻¹	3.26[1.90:	4.62]	2.8×10 ⁻⁶	-0.09[-0.83:	0.64]	8.0×10 ⁻¹	-0.01[-0.61:	0.59]	9.8×10 ⁻¹	-0.34[-0.89:	0.20]	2.2×10 ⁻¹	-0.17[-0.52:	0.18]	3.5×10 ⁻¹		
M-LDL-PL	Phospholipids in medium LDL	6.74[4.74:	8.74]	8.3×10 ⁻¹¹	2.52[-0.11:	5.15]	6.1×10 ⁻²	0.26[-2.33:	2.86]	8.4×10 ⁻¹	3.84[2.49:	5.20]	2.8×10 ⁻⁸	-0.13[-0.87:	0.60]	7.2×10 ⁻¹	0.12[-0.49:	0.72]	7.1×10 ⁻¹	-0.28[-0.82:	0.27]	3.2×10 ⁻¹	-0.11[-0.46:	0.25]	5.6×10 ⁻¹		
M-LDL-C	Total cholesterol in medium LDL	5.56[3.55:	7.57]	8.4×10 ⁻⁸	2.46[-0.19:	5.10]	6.9×10 ⁻²	0.24[-2.36:	2.85]	8.5×10 ⁻¹	3.28[1.91:	4.64]	2.5×10 ⁻⁶	-0.01[-0.75:	0.73]	9.8×10 ⁻¹	0.03[-0.57:	0.63]	9.1×10 ⁻¹	-0.33[-0.87:	0.21]	2.3×10 ⁻¹	-0.13[-0.48:	0.22]	4.7×10 ⁻¹		
M-LDL-CE	Cholesterol esters in medium LDL	5.00[2.98:	7.02]	1.5×10 ⁻⁶	2.45[-0.20:	5.09]	7.0×10 ⁻²	0.14[-2.47:	2.75]	9.2×10 ⁻¹	2.98[1.62:	4.35]	1.9×10 ⁻⁵	0.01[-0.73:	0.75]	9.8×10 ⁻¹	-0.04[-0.64:	0.56]	9.0×10 ⁻¹	-0.35[-0.89:	0.20]	2.1×10 ⁻¹	-0.16[-0.51:	0.20]	3.8×10 ⁻¹		
M-LDL-FC	Free cholesterol in medium LDL	7.09[5.11:	9.08]	6.3×10 ⁻¹²	2.33[-0.31:	4.97]	8.3×10 ⁻²	0.41[-2.18:	3.01]	7.5×10 ⁻¹	4.03[2.67:	5.38]	5.6×10 ⁻⁹	-0.06[-0.80:	0.67]	8.6×10 ⁻¹	0.28[-0.32:	0.88]	3.6×10 ⁻¹	-0.26[-0.79:	0.28]	3.5×10 ⁻¹	-0.03[-0.38:	0.32]	8.8×10 ⁻¹		
M-LDL-TG	Triglycerides in medium LDL	-1.62[-3.69:	0.46]	1.3×10 ⁻¹	2.58[-0.07:	5.22]	5.6×10 ⁻²	-1.88[-4.52:	0.75]	1.6×10 ⁻¹	-0.54[-1.92:	0.85]	4.5×10 ⁻¹	-0.68[-1.42:	0.06]	7.0×10 ⁻²	-1.16[-1.76:-0.55]	1.9×10 ⁻⁴	-0.53[-1.06:	0.01]	5.4×10 ⁻²	-0.78[-1.13:-0.42]	1.6×10 ⁻⁵					
S-LDL-P	Concentration of small LDL particles	5.24[3.22:	7.27]	5.0×10 ⁻⁷	2.51[-0.14:	5.15]	6.3×10 ⁻²	-0.13[-2.74:	2.48]	9.2×10 ⁻¹	3.03[1.66:	4.40]	1.4×10 ⁻⁵	0.10[-0.64:	0.84]	7.9×10 ⁻¹	0.03[-0.57:	0.64]	9.1×10 ⁻¹	-0.32[-0.86:	0.23]	2.5×10 ⁻¹	-0.10[-0.45:	0.26]	5.9×10 ⁻¹		
S-LDL-L	Total lipids in small LDL	5.73[3.72:	7.75]	3.6×10 ⁻⁸	2.45[-0.19:	5.09]	6.9×10 ⁻²	0.14[-2.47:	2.74]	9.2×10 ⁻¹	3.32[1.96:	4.69]	1.8×10 ⁻⁶	0.13[-0.61:	0.86]	7.4×10 ⁻¹	0.12[-0.48:	0.72]	6.9×10 ⁻¹	-0.32[-0.86:	0.22]	2.5×10 ⁻¹	-0.07[-0.42:	0.29]	7.2×10 ⁻¹		
S-LDL-PL	Phospholipids in small LDL	7.12[5.12:	9.12]	1.0×10 ⁻¹²	2.22[-0.42:	4.85]	1.0×10 ⁻¹	0.07[-2.53:	2.66]	9.6×10 ⁻¹	3.89[2.53:	5.25]	2.0×10 ⁻⁸	0.30[-0.44:	1.03]	4.3×10 ⁻¹	0.38[-0.22:	0.98]	2.2×10 ⁻¹	-0.12[-0.66:	0.43]	6.8×10 ⁻¹	0.15[-0.20:	0.51]	4.0×10 ⁻¹		
S-LDL-C	Total cholesterol in small LDL	5.59[3.58:	7.60]	7.3×10 ⁻⁸	2.43[-0.22:	5.08]	7.2×10 ⁻²	0.39[-2.22:	3.01]	7.7×10 ⁻¹	3.33[1.96:	4.70]	1.8×10 ⁻⁶	0.10[-0.64:	0.84]	8.0×10 ⁻¹	0.15[-0.45:	0.75]	6.3×10 ⁻¹	-0.33[-0.87:	0.21]	2.3×10 ⁻¹	-0.07[-0.42:	0.28]	7.0×10 ⁻¹		
S-LDL-CE	Cholesterol esters in small LDL	5.09[3.07:	7.11]	1.0×10 ⁻⁶	2.48[-0.17:	5.13]	6.7×10 ⁻²	0.40[-2.22:	3.02]	7.6×10 ⁻¹	3.11[1.74:	4.48]	8.5×10 ⁻⁶	0.08[-0.66:	0.82]	8.3×10 ⁻¹	0.09[-0.51:	0.69]	7.7×10 ⁻¹	-0.33[-0.87:	0.21]	2.3×10 ⁻¹	-0.09[-0.45:	0.26]	6.1×10 ⁻¹		
S-LDL-FC	Free cholesterol in small LDL	6.77[4.78:	8.76]	6.1×10 ⁻¹¹	2.07[-0.58:	4.71]	1.3×10 ⁻¹	0.15[-2.45:	2.75]	9.1×10 ⁻¹	3.72[2.36:	5.08]	7.7×10 ⁻⁸	0.15[-0.59:	0.89]	7.0×10 ⁻¹	0.34[-0.26:	0.94]	2.7×10 ⁻¹	-0.32[-0.86:	0.22]	2.5×10 ⁻¹	0.01[-0.34:	0.37]	9.4×10 ⁻¹		
S-LDL-TG	Triglycerides in small LDL	-1.16[-3.24:	0.91]	2.7×10 ⁻¹	2.68[-0.02:	5.37]	5.1×10 ⁻²	-2.32[-4.92:	0.28]	8.2×10 ⁻²	-0.47[-1.86:	0.92]	5.1×10 ⁻¹	-0.31[-1.04:	0.42]	4.1×10 ⁻¹	-1.07[-1.68:-0.46]	6.4×10 ⁻⁴	-0.57[-1.12:-0.02]	4.1×10 ⁻²	-0.68[-1.04:-0.32]	1.9×10 ⁻⁴						
XL-HDL-P	Concentration of very large HDL particles	2.80[0.69:	4.91]	9.6×10 ⁻³	1.15[-1.88:	4.18]	4.6×10 ⁻¹	3.47[0.80:	6.14]	1.2×10 ⁻²	2.62[1.17:	4.07]	4.1×10 ⁻⁴	1.30[0.46:	2.14]	2.4×10 ⁻³	1.88[1.19:	2.56]	8.8×10 ⁻⁸	0.58[-0.05:	1.22]	7.3×10 ⁻²		
XL-HDL-L	Total lipids in very large HDL	3.44[1.33:	5.54]	1.4×10 ⁻³	1.24[-1.75:	4.23]	4.2×10 ⁻¹	3.60[0.93:	6.27]	9.1×10 ⁻³	2.97[1.52:	4.42]	5.7×10 ⁻⁵	1.42[0.58:	2.26]	9.4×10 ⁻⁴	1.91[1.23:	2.60]	5.6×10 ⁻⁸	0.65[0.02:	1.29]		
XL-HDL-PL	Phospholipids in very large HDL	2.05[-0.07:	4.18]	5.9×10 ⁻²	1.20[-1.85:	4.26]	4.4×10 ⁻¹	3.69[1.04:	6.34]	7.1×10 ⁻³	2.35[0.90:	3.81]	1.6×10 ⁻³	1.56[0.71:	2.40]	3.1×10 ⁻⁴	1.97[1.27:	2.67]	4.2×10 ⁻⁸	0.69[0.03:	1.35]		
XL-HDL-C	Total cholesterol in very large HDL	4.18[2.10:	6.26]	9.0×10 ⁻⁵	1.12[-1.84:	4.07]	4.6×10 ⁻¹	3.20[0.53:	5.86]	2.0×10 ⁻²	3.18[1.74:	4.61]	1.4×10 ⁻⁵	1.21[0.38:	2.04]	4.2×10 ⁻³	1.82[1.16:	2.49]	9.6×10 ⁻⁸	0.53[-0.08:	1.14]	8.7×10 ⁻²		
XL-HDL-CE	Cholesterol esters in very large HDL	4.22[2.15:	6.30]	7.4×10 ⁻⁵	1.11[-1.83:	4.05]	4.6×10 ⁻¹	3.03[0.38:	5.68]	2.6×10 ⁻²	3.14[1.71:	4.57]	1.6×10 ⁻⁵	1.17[0.35:	1.99]	5.3×10 ⁻³	1.70[1.04:	2.36]	4.6×10 ⁻⁷	0.44[-0.16:	1.04]	1.5×10 ⁻¹		
XL-HDL-FC	Free cholesterol in very large HDL	3.60[1.50:	5.70]	8.3×10 ⁻⁴	1.06[-1.96:	4.07]	4.9×10 ⁻¹	2.77[0.36:	2.95]	5.0×10 ⁻³	1.21[0.36:	2.05]	5.0×10 ⁻³	1.21[0.38:	2.05]	5.0×10 ⁻³	0.70[-0.08:	1.32]	2.8×10 ⁻²	0.88[0.38:	1.38]	5.8×10 ⁻⁴		
XL-HDL-TG	Triglycerides in very large HDL	2.78[0.69:	4.87]	9.3×10 ⁻³	1.55[-1.19:	4.30]	2.7×10 ⁻¹	-0.71[-3.31:	1.88]	5.9×10 ⁻¹	1.44[0.04:	2.84]	4.3×10 ⁻²	-0.22[-1.01:	0.57]	5.9×10 ⁻¹	-0.35[-0.95:	0.25]	2.5×10 ⁻¹	-0.19[-0.72:	0.35]	4.9×10 ⁻¹	-0.25[-0.61:	0.10]	1.7×10 ⁻¹		
L-HDL-P	Concentration of large HDL particles	1.65[-0.49:	3.79]	1.3×10 ⁻¹	0.03[-3.36:	3.43]	9.8×10 ⁻¹	2.62[-0.06:	5.31]	5.8×10 ⁻²	1.64[0.14:	3.14]	3.2×10 ⁻²	1.28[0.44:	2.11]	2.8×10 ⁻³	1.75[1.05:	2.45]	1.1×10 ⁻⁶	0.35[-0.30:	0.99]	2.9×10 ⁻¹	1.06[0.65:	1.47]
L-HDL-L	Total lipids in large HDL	2.21[0.08:	4.34]	4.3×10 ⁻²	0.35[-3.10:	3.80]	8.4×10 ⁻¹	2.64[-0.04:	5.31]	5.5×10 ⁻²	1.99[0.49:	3.49]	9.3×10 ⁻³	1.43[0.59:	2.27]	9.0×10 ⁻⁴	1.81[1.11:	2.51]	4.8×10 ⁻⁷	0.39[-0.26:	1.04]	2.4×10 ⁻¹	1.14[0.72:	1.55]
L-HDL-PL	Phospholipids in large HDL	2.53[0.40:	4.66]	2.0×10 ⁻²	0.41[-3.00:	3.82]	8.1×10 ⁻¹	2.62[-0.05:	5.30]	5.6×10 ⁻²	1.75[0.63:	2.80]	5.8×10 ⁻⁴	1.47[0.63:	2.30]	5.8×10 ⁻⁴	1.75[1.05:	2.44]	9.9×10 ⁻⁷	0.37[-0.27:	1.01]	2.5×10 ⁻¹	1.11[0.70:	1.52]
L-HDL-C	Total cholesterol in large HDL	1.93[-0.19:	4.05]	7.5×10 ⁻²	-0.16[-3.56:	3.24]	9.3×10 ⁻¹	2.84[0.17:	5.50]	3.9×10 ⁻²	1.81[0.32:	3.30]	1.8×10 ⁻²	1.34[0.50:	2.18]	1.8×10 ⁻³	1.90[1.20:	2.61]	1.3×10 ⁻⁷	0.50[-0.16:	1.16]	1.4×10 ⁻¹		
L-HDL-CE	Cholesterol esters in large HDL	1.99[-0.13:	4.12]	6.6×10 ⁻²	-0.20[-3.60:	3.20]	9.1×10 ⁻¹	2.83[0.16:	5.50]	3.9×10 ⁻²	1.83[0.34:	3.33]	1.6×10 ⁻²	1.33[0.49:	2.17]	1.9×10 ⁻³	1.89[1.19:	2.59]	1.5×10 ⁻⁷	0.51[-0.15:	1.17]	1.3×10 ⁻¹		
L-HDL-FC	Free cholesterol in large HDL	1.25[-0.87:	3.25]	2.5×10 ⁻¹	-0.31[-3.65:	3.04]	8.6×10 ⁻¹	2.72[0.05:	5.39]	4.8×10 ⁻²	1.40[-0.09:	2.89]	6.6×10 ⁻²	1.24[0.40:	2.09]	3.9×10 ⁻³	1.88[1.17:	2.58]	2.1×10 ⁻⁷	0.44[-0.22:	1.10]	1.9×10 ⁻¹		
L-HDL-TG	Triglycerides in large HDL	-3.64[-5.79:	-1.49]	9.7×10 ⁻⁴	1.76[-1.46:	4.97]	2.8×10 ⁻¹	-0.04[-2.71:	2.63]	9.8×10 ⁻¹	-1.37[-2.86:	0.11]	7.0×10 ⁻²	0.19[-0.62:	1.01]	6.4×10 ⁻¹	0.05[-0.60:	0.70]	8.8×10 ⁻¹	-0.71[-1.30:-0.13]	1.6×10 ⁻²	-0.25[-0.63:	0.13]	2.0×10 ⁻¹				
M-HDL-P	Concentration of medium HDL particles	5.37[3.28:	7.46]	6.3×10 ⁻⁷	0.46[-2.56:	3.48]	7.7×10 ⁻¹	2.30[-0.31:	4.90]	8.6×10 ⁻²	3.33[1.89:	4.76]	5.4×10 ⁻⁶	1.12[0.37:	1.87]	3.5×10 ⁻³	1.11[0.48:	1.74]	5.8×10 ⁻⁴	0.33[-0.21:	0.86]	2.3×10 ⁻¹	0.76[0.40:	1.12]
M-HDL-L	Total lipids in medium HDL	5.42[3.28:	7.56]	4.4×10 ⁻⁷	0.52[-2.51:	3.54]	7.4×10 ⁻¹	2.07[-0.61:	4.75]	1.3×10 ⁻¹	3.33[1.88:	4.77]	6.3×10 ⁻⁶	1.17[0.42:	1.93]	2.3×10 ⁻³	1.17[0.54:	1.81]	2.9×10 ⁻⁴	0.38[-0.16:	0.92]	1.7×10 ⁻¹	0.82[0.46:	1.18]
M-HDL-PL	Phospholipids in medium HDL	5.09[2.99:	7.19]	2.4×10 ⁻⁶	0.81[-2.24:	3.87]	6.0×10 ⁻¹	2.16[-0.44:	4.77]	1.1×10 ⁻¹	3.24[1.80:</																

VLDL-TG	Triglycerides in VLDL	-0.11[-2.18: 1.97]	9.2×10 ⁻¹	2.61[-0.38: 5.61]	8.7×10 ⁻²	-2.71[-5.29:-0.13]	4.1×10 ⁻²	-0.29[-1.71: 1.14]	6.9×10 ⁻¹	-0.89[-1.68:-0.11]	2.5×10 ⁻²	-1.38[-2.04:-0.72]	4.0×10 ⁻⁵	-0.70[-1.31:-0.08]	2.6×10 ⁻²	-0.99[-1.37:-0.60]	6.8×10 ⁻⁷
LDL-TG	Triglycerides in LDL	-2.34[-4.42: -0.25]	2.9×10 ⁻²	2.66[0.03: 5.28]	4.7×10 ⁻²	-2.53[-5.16: 0.09]	6.1×10 ⁻²	-1.00[-2.38: 0.39]	1.6×10 ⁻¹	-0.62[-1.36: 0.12]	1.0×10 ⁻¹	-1.18[-1.79:-0.57]	1.6×10 ⁻⁴	-0.69[-1.24:-0.15]	1.2×10 ⁻²	-0.84[-1.20:-0.49]	3.5×10 ⁻⁶
HDL-TG	Triglycerides in HDL	-3.66[-5.78: -1.55]	7.2×10 ⁻⁴	2.84[0.06: 5.61]	4.5×10 ⁻²	-2.85[-5.41:-0.29]	3.1×10 ⁻²	-1.75[-3.15:-0.34]	1.5×10 ⁻²	-0.51[-1.25: 0.24]	1.8×10 ⁻¹	-1.06[-1.67:-0.45]	6.4×10 ⁻⁴	-0.68[-1.21:-0.15]	1.3×10 ⁻²	-0.77[-1.12:-0.42]	2.0×10 ⁻⁵
TotPG	Total phosphoglycerides	2.49[0.37: 4.61]	2.2×10 ⁻²	2.04[-0.84: 4.92]	1.7×10 ⁻¹	-0.58[-3.29: 2.13]	6.8×10 ⁻¹	1.50[0.06: 2.95]	4.1×10 ⁻²	0.28[-0.50: 1.06]	4.8×10 ⁻¹	0.33[-0.29: 0.96]	3.0×10 ⁻¹	-0.20[-0.73: 0.32]	4.5×10 ⁻¹	0.07[-0.28: 0.43]	6.8×10 ⁻¹
TG/PG	Ratio of triglycerides to phosphoglycerides	-2.70[-4.75: -0.64]	1.0×10 ⁻²	1.45[-1.67: 4.57]	3.6×10 ⁻¹	-3.28[-5.89:-0.67]	1.5×10 ⁻²	-2.00[-3.43:-0.57]	6.3×10 ⁻³	-1.27[-2.06:-0.47]	1.7×10 ⁻³	-2.17[-2.85:-1.48]	5.7×10 ⁻¹⁰	-0.62[-1.22:-0.01]	4.5×10 ⁻²	-1.29[-1.69:-0.90]	1.1×10 ⁻¹⁰
PC	Phosphatidylcholine and other choline	2.32[0.20: 4.44]	3.3×10 ⁻²	1.89[-1.03: 4.82]	2.0×10 ⁻¹	-0.34[-3.10: 2.42]	8.1×10 ⁻¹	1.47[0.01: 2.93]	4.8×10 ⁻²	0.27[-0.51: 1.05]	5.0×10 ⁻¹	0.58[-0.05: 1.22]	7.1×10 ⁻²	-0.20[-0.73: 0.33]	4.6×10 ⁻¹	0.15[-0.21: 0.51]	4.0×10 ⁻¹
SM	Sphingomyelins	2.61[0.55: 4.68]	1.3×10 ⁻²	1.99[-0.88: 4.86]	1.8×10 ⁻¹	0.50[-2.28: 3.29]	7.2×10 ⁻¹	1.90[0.46: 3.33]	9.7×10 ⁻³	0.21[-0.56: 0.99]	5.9×10 ⁻¹	0.65[0.01: 1.30]	4.7×10 ⁻²	-0.19[-0.72: 0.34]	4.9×10 ⁻¹	0.17[-0.20: 0.53]	3.7×10 ⁻¹
TotCho	Total choline	1.51[-0.61: 3.62]	1.6×10 ⁻¹	2.06[-0.81: 4.94]	1.6×10 ⁻¹	0.15[-2.64: 2.93]	9.2×10 ⁻¹	1.28[-0.17: 2.73]	8.5×10 ⁻²	0.33[-0.45: 1.11]	4.1×10 ⁻¹	0.70[0.07: 1.34]	3.1×10 ⁻²	-0.13[-0.66: 0.40]	6.3×10 ⁻¹	0.24[-0.13: 0.60]	2.0×10 ⁻¹
ApoA1	Apolipoprotein A-I	5.69[3.58: 7.79]	1.6×10 ⁻⁷	1.25[-1.83: 4.32]	4.3×10 ⁻¹	1.90[-0.92: 4.72]	1.9×10 ⁻¹	3.62[2.14: 5.10]	1.6×10 ⁻⁶	1.11[0.29: 1.93]	7.8×10 ⁻³	1.69[1.02: 2.36]	8.5×10 ⁻⁷	0.05[-0.51: 0.61]	8.6×10 ⁻¹	0.81[0.43: 1.19]	3.1×10 ⁻⁵
ApoB	Apolipoprotein B	3.25[1.21: 5.29]	1.9×10 ⁻³	2.81[0.17: 5.45]	3.7×10 ⁻²	-1.68[-4.24: 0.88]	2.0×10 ⁻¹	1.73[0.37: 3.10]	1.3×10 ⁻²	-0.41[-1.14: 0.33]	2.8×10 ⁻¹	-0.64[-1.25:-0.02]	4.2×10 ⁻²	-0.63[-1.21:-0.05]	3.3×10 ⁻²	-0.58[-0.94:-0.21]	1.9×10 ⁻³
ApoB/ApoA1	Ratio of apolipoprotein B to apolipoprotein A-I	0.56[-1.48: 2.61]	5.9×10 ⁻¹	2.76[-0.06: 5.59]	5.5×10 ⁻²	-2.26[-4.83: 0.31]	8.7×10 ⁻²	0.27[-1.12: 1.66]	7.0×10 ⁻¹	-0.95[-1.69:-0.20]	1.3×10 ⁻²	-1.40[-2.05:-0.75]	2.8×10 ⁻⁵	-0.69[-1.32:-0.07]	3.0×10 ⁻²	-1.01[-1.40:-0.62]	3.0×10 ⁻⁷
TotFA	Total fatty acids	1.59[-0.41: 3.69]	1.4×10 ⁻¹	2.73[-0.05: 5.50]	5.4×10 ⁻²	-2.26[-4.90: 0.38]	9.5×10 ⁻²	0.78[-0.63: 2.20]	2.8×10 ⁻¹	-0.31[-1.06: 0.44]	4.2×10 ⁻¹	-0.59[-1.19: 0.02]	5.9×10 ⁻²	-0.32[-0.86: 0.23]	2.6×10 ⁻¹	-0.41[-0.76:-0.05]	2.5×10 ⁻²
UnSat	Estimated degree of unsaturation	-1.96[-4.02: 0.09]	6.1×10 ⁻²	-0.74[-3.73: 2.26]	6.3×10 ⁻¹	0.82[-1.82: 3.46]	5.5×10 ⁻¹	-0.88[-2.30: 0.55]	2.3×10 ⁻¹	1.83[1.08: 2.59]	1.9×10 ⁻⁶	0.60[-0.02: 1.23]	5.9×10 ⁻²	-0.55[-1.08:-0.01]	4.4×10 ⁻²	0.36[0.00: 0.72]	4.7×10 ⁻²
DHA	22:6, docosahexaenoic acid	-0.08[-2.14: 1.98]	9.4×10 ⁻¹	2.53[-0.49: 5.54]	1.0×10 ⁻¹	-1.78[-4.36: 0.80]	1.8×10 ⁻¹	-0.02[-1.44: 1.40]	9.8×10 ⁻¹	1.42[0.66: 2.17]	2.4×10 ⁻⁴	-0.32[-0.93: 0.28]	3.0×10 ⁻¹	-0.51[-1.03: 0.02]	6.1×10 ⁻²	-0.03[-0.38: 0.33]	8.9×10 ⁻¹
LA	18:2, linoleic acid	2.53[0.44: 4.62]	3.8×10 ⁻²	2.87[0.16: 5.57]	3.8×10 ⁻²	-0.96[-3.69: 1.77]	4.9×10 ⁻¹	1.69[0.27: 3.10]	1.9×10 ⁻²	-0.49[-1.23: 0.26]	2.0×10 ⁻¹	0.11[-0.50: 0.73]	7.2×10 ⁻¹	-0.50[-1.03: 0.03]	6.4×10 ⁻²	-0.29[-0.65: 0.06]	1.0×10 ⁻¹
FAw3	Omega-3 fatty acids	1.60[-0.48: 3.68]	1.3×10 ⁻¹	3.01[0.10: 5.91]	4.2×10 ⁻²	-1.75[-4.34: 0.84]	1.9×10 ⁻¹	0.93[-0.49: 2.35]	2.0×10 ⁻¹	1.11[0.36: 1.86]	3.7×10 ⁻³	-0.33[-0.94: 0.28]	2.9×10 ⁻¹	-0.62[-1.14:-0.09]	2.3×10 ⁻²	-0.14[-0.49: 0.21]	4.4×10 ⁻¹
FAw6	Omega-6 fatty acids	2.76[0.67: 4.85]	1.0×10 ⁻²	2.75[0.00: 5.50]	5.0×10 ⁻²	-0.93[-3.69: 1.82]	5.1×10 ⁻¹	1.77[0.35: 3.20]	1.5×10 ⁻²	-0.12[-0.88: 0.63]	7.5×10 ⁻¹	0.18[-0.43: 0.80]	5.6×10 ⁻¹	-0.42[-0.95: 0.10]	1.2×10 ⁻¹	-0.16[-0.51: 0.20]	3.8×10 ⁻¹
PUFA	Polynunsaturated fatty acids	2.62[0.53: 4.72]	1.4×10 ⁻²	2.85[0.08: 5.62]	4.4×10 ⁻²	-1.25[-3.99: 1.49]	3.7×10 ⁻¹	1.64[0.21: 3.06]	2.5×10 ⁻²	0.10[-0.65: 0.86]	7.9×10 ⁻¹	0.10[-0.51: 0.71]	7.4×10 ⁻¹	-0.45[-0.98: 0.07]	9.2×10 ⁻²	-0.15[-0.50: 0.21]	4.2×10 ⁻¹
MUFA	Monounsaturated fatty acids; 16:1, 18:1	-0.76[-2.84: 1.32]	4.8×10 ⁻¹	2.51[-0.29: 5.31]	7.9×10 ⁻²	-2.96[-5.55:-0.36]	2.7×10 ⁻²	-0.58[-1.98: 0.82]	4.2×10 ⁻¹	-0.78[-1.53:-0.03]	4.2×10 ⁻²	-1.24[-1.86:-0.62]	8.4×10 ⁻⁵	-0.46[-1.02: 0.10]	1.1×10 ⁻¹	-0.81[-1.17:-0.44]	1.4×10 ⁻⁵
SFA	Saturated fatty acids	2.65[0.55: 4.74]	1.3×10 ⁻²	2.49[-0.28: 5.26]	7.8×10 ⁻²	-1.63[-4.27: 1.01]	2.3×10 ⁻¹	1.38[-0.03: 2.80]	5.5×10 ⁻²	-0.26[-1.01: 0.49]	5.0×10 ⁻¹	-0.41[-1.01: 0.20]	1.9×10 ⁻¹	-0.01[-0.55: 0.54]	9.9×10 ⁻¹	-0.20[-0.56: 0.16]	2.7×10 ⁻¹
Glc	Glucose	5.16[3.15: 7.17]	6.4×10 ⁻⁷	1.27[-1.59: 4.13]	3.8×10 ⁻¹	-2.85[-5.55:-0.16]	4.0×10 ⁻²	2.05[0.64: 3.45]	4.2×10 ⁻³	0.08[-0.67: 0.83]	8.4×10 ⁻¹	-0.51[-1.20: 0.18]	1.5×10 ⁻¹	-0.48[-1.03: 0.08]	9.4×10 ⁻²	-0.35[-0.72: 0.03]	6.9×10 ⁻²
Lac	Lactate	-0.30[-2.37: 1.77]	7.8×10 ⁻¹	2.33[-0.54: 5.20]	1.1×10 ⁻¹	-0.10[-2.83: 2.63]	9.4×10 ⁻¹	0.41[-1.02: 1.84]	5.7×10 ⁻¹	-0.22[-0.98: 0.54]	5.7×10 ⁻¹	-1.80[-2.43:-1.16]	3.9×10 ⁻⁸	-0.37[-0.91: 0.16]	1.7×10 ⁻¹	-0.79[-1.15:-0.43]	1.6×10 ⁻⁵
Pyr	Pyruvate	4.82[2.76: 6.88]	5.5×10 ⁻⁶	2.83[-0.13: 5.80]	6.1×10 ⁻²	1.22[-1.55: 4.00]	3.9×10 ⁻¹	3.37[1.93: 4.82]	4.7×10 ⁻⁶	0.27[-0.48: 1.03]	4.8×10 ⁻¹	-1.73[-2.36:-1.10]	9.4×10 ⁻⁸	-0.45[-0.98: 0.09]	1.0×10 ⁻¹	-0.70[-1.06:-0.34]	1.3×10 ⁻⁴
Cit	Citrate	-4.97[-7.01: -2.93]	2.2×10 ⁻⁶	-2.62[-5.47: 0.23]	7.1×10 ⁻²	-0.74[-3.66: 2.18]	6.2×10 ⁻²	-3.34[-4.78:-1.90]	5.7×10 ⁻⁶	-2.43[-3.16:-1.69]	1.1×10 ⁻³	-1.87[-2.51:-1.22]	1.7×10 ⁻⁸	-1.50[-2.01:-0.98]	1.9×10 ⁻⁸	-1.82[-2.18:-1.47]	7.1×10 ⁻¹⁶
GloI	Glycerol	-11.03[-13.96: -8.10]	1.7×10 ⁻¹²	-1.53[-4.70: 1.64]	3.4×10 ⁻¹	-3.09[-6.36: 0.17]	6.6×10 ⁻²	-5.57[-7.37:-3.77]	1.2×10 ⁻⁹	-1.31[-2.10:-0.53]	3.9×10 ⁻¹¹	-1.46[-2.20:-0.72]	1.1×10 ⁻⁴	-1.77[-2.19:-1.34]	1.1×10 ⁻⁴	-1.87[-2.18:-1.47]	7.5×10 ⁻²⁴
Ala	Alanine	7.14[5.11: 9.17]	1.2×10 ⁻¹¹	2.07[-0.73: 4.87]	1.5×10 ⁻¹	-2.28[-4.97: 0.42]	1.0×10 ⁻¹	3.32[1.92: 4.72]	3.5×10 ⁻⁶	-1.18[-1.92:-0.43]	2.0×10 ⁻³	-1.92[-2.54:-1.29]	2.4×10 ⁻⁹	-1.11[-1.65:-0.58]	4.6×10 ⁻⁵	-1.39[-1.75:-1.03]	2.4×10 ⁻¹⁴
Gln	Glutamine	-7.87[-12.31: -3.43]	7.2×10 ⁻⁴	0.32[-2.77: 3.41]	8.4×10 ⁻¹	-1.41[-4.17: 1.36]	3.2×10 ⁻¹	-1.92[-3.79:-0.05]	4.4×10 ⁻²	-0.16[-0.95: 0.62]	6.8×10 ⁻¹	0.53[-0.11: 1.16]	1.0×10 ⁻¹	-0.61[-1.17:-0.06]	3.1×10 ⁻²	-0.13[-0.50: 0.24]	5.0×10 ⁻¹
Gly	Glycine	-12.25[-14.10:-10.41]	1.4×10 ⁻³⁴	-2.04[-4.91: 0.83]	1.6×10 ⁻¹	-5.53[-8.30:-2.77]	1.4×10 ⁻¹	-8.37[-9.73:-7.02]	7.3×10 ⁻³⁴	-1.27[-2.02:-0.52]	9.6×10 ⁻⁴	-1.49[-2.18:-0.81]	1.9×10 ⁻⁵	-1.18[-1.72:-0.64]	1.9×10 ⁻⁵	-1.29[-1.66:-0.92]	6.3×10 ⁻¹²
His	Histidine	1.24[-0.82: 3.30]	2.4×10 ⁻¹	2.18[-0.66: 5.03]	1.3×10 ⁻¹	0.74[-2.02: 3.50]	6.0×10 ⁻¹	1.35[-0.58: 2.77]	6.5×10 ⁻²	0.20[-0.95: 0.94]	6.1×10 ⁻¹	-0.31[-0.93: 0.31]	3.3×10 ⁻¹	-0.98[-1.50:-0.46]	2.3×10 ⁻⁴	-0.51[-0.86:-0.15]	5.0×10 ⁻³
Ile	Isoleucine	2.20[0.13: 4.27]	3.7×10 ⁻²	0.90[-2.10: 3.90]	5.6×10 ⁻¹	-3.11[-5.86:-0.36]	2.8×10 ⁻²	0.43[-1.02: 1.88]	5.6×10 ⁻¹	-1.37[-2.15:-0.59]	5.4×10 ⁻⁴	-2.01[-2.74:-1.27]	9.9×10 ⁻⁸	-1.44[-2.08:-0.80]	1.1×10 ⁻⁵	-1.60[-2.01:-1.19]	2.2×10 ⁻¹⁴
Leu	Leucine	5.37[3.34: 7.41]	3.0×10 ⁻⁷	0.81[-2.17: 3.78]	6.0×10 ⁻¹	-2.95[-5.71:-0.18]	3.9×10 ⁻²	2.07[0.63: 3.51]	4.7×10 ⁻³	-1.17[-1.95:-0.40]	3.1×10 ⁻³	-1.34[-2.09:-0.59]	4.5×10 ⁻⁴	-1.45[-2.09:-0.80]	1.1×10 ⁻⁵	-1.34[-1.75:-0.92]	2.2×10 ⁻¹⁰
Val	Valine	8.79[6.83: 10.75]	1.2×10 ⁻¹⁷	1.13[-1.81: 4.08]	4.5×10 ⁻¹	-1.03[-3.76: 1.70]	4.6×10 ⁻¹	-0.87[-1.63:-0.11]	2.5×10 ⁻²	-0.65[-1.37: 0.07]	7.6×10 ⁻²	-1.64[-2.37: -0.91]	4.4×10 ⁻⁴	-0.88[-1.27:-0.48]	1.4×10 ⁻⁵	-1.05[-1.44:-0.66]	1.4×10 ⁻⁵
Phe	Phenylalanine	-11.70[-13.54: -9.86]	4.4×10 ⁻³²	-1.58[-4.51: 1.36]	2.9×10 ⁻¹	-5.17[-7.86:-2.47]	2.5×10 ⁻⁴	-7.92[-9.27:-6.57]	1.2×10 ⁻³⁰	-2.04[-2.80:-1.28]	1.4×10 ⁻⁷	-2.11[-2.77:-1.45]	5.5×10 ⁻¹⁰	-1.18[-1.76:-0.61]	5.1×10 ⁻⁵	-1.69[-2.07:-1.32]	1.1×10 ⁻¹⁸
Tyr	Tyrosine	5.72[3.70: 7.75]	4.4×10 ⁻⁸	-0.16[-2.98: 2.67]	9.1×10 ⁻¹	2.11[-0.60: 4.83]	1.3×10 ⁻¹	3.29[1.89: 4.70]	4.5×10 ⁻⁶	-0.16[-0.91: 0.59]	6.8×10 ⁻¹	0.15[-0.52: 0.81]	6.7×10 ⁻¹	-0.47[-1.04: 0.10]	1.0×10 ⁻¹	-0.20[-0.57: 0.18]	3.0×10 ⁻¹
Ace	Acetate	0.71[-1.38: 2.79]	5.1×10 ⁻¹	0.35[-2.37: 3.07]	8.0×10 ⁻¹	-0.22[-2.99: 2.55]	8.8×10 ⁻¹	0.36[-1.06: 1.79]	6.2×10 ⁻¹	0.28[-0.43: 1.00]	4.4×10 ⁻¹	0.52[-0.11: 1.14]	1.1×10 ⁻¹	-0.37[-0.85: 0.19]	2.2×10 ⁻¹	0.08[-0.27: 0.43]	6.5×10 ⁻¹
AcAce	Acetoacetate	1.48[-0.59: 3.54]	1.6×10 ⁻¹	0.16[-2.65: 2.98]	9.1×10 ⁻¹	-1.42[-4.18: 1.35]	3.2×10 ⁻¹	0.37[-1.06: 1.80]	6.1×10 ⁻¹	0.90[0.15: 1.64]	1.8×10 ⁻²	0.03[-0.60: 0.66]	9.2×10 ⁻¹	-0.30[-0.82: 0.22]	2.6×10 ⁻¹	0.07[-0.28: 0.43]	6.8×10 ⁻¹
boHBut	3-hydroxybutyrate	-0.45[-2.54: 1.64]	6.7×10 ⁻¹	0.06[-2.80: 2.92]	9.7×10 ⁻¹	0.10[-2.70: 2.90]	9.5×10 ⁻¹	-0.17[-1.62: 1.27]	8.1×10 ⁻¹	0.40[-0.33: 1.14]	2.9×10 ⁻¹	0.17[-0.46: 0.81]	5.9×10 ⁻¹	-0.04[-0.74: 0.66]	9.1×10 ⁻¹	0.17[-0.23: 0.57]	4.0×10 ⁻¹
Alb	Albumin	6.57[4.55: 8.58]	3.3×10 ⁻¹⁰	-0.01[-2.88: 2.85]	9.9×10 ⁻¹	-1.86[-4.41: 0.70]	1.6×10 ⁻¹	2.55[1.16: 3.93]	3.1×10 ⁻⁴	-0.16[-0.94: 0.61]	6.8×10 ⁻¹	-0.59[-1.20: 0.02]	5.8×10 ⁻²	-1.03[-1.55:-0.51]	1.1×10 ⁻⁴	-0.70[-1.06:-0.35]	9.4×10 ⁻⁵
Gp	Glycoprotein acetyls, mainly a1-acid glycoprotein	-3.63[-5.72: -1.53]	7.4×10 ⁻⁴	0.37[-2.66: 3.41]	8.1×10 ⁻¹	-3.88[-6.58:-1.19]	5.4×10 ⁻¹	-2.79[-4.24:-1.33]	1.7×10 ⁻⁴	-0.93[-1.72:-0.14]	2.1×10 ⁻²	-1.82[-2.49:-1.15]	9.7×10 ⁻⁸	-0.48[-1.06			

M-VLDL-C-%	Total cholesterol to total lipids ratio in medium VLDL	-0.82[-2.91: 1.27]	4.4×10 ⁻¹	1.28[-1.57: 4.14]	3.8×10 ⁻¹	1.50[-1.12: 4.12]	2.6×10 ⁻¹	0.38[-1.04: 1.80]	6.0×10 ⁻¹	-0.17[-0.92: 0.57]	6.5×10 ⁻¹	0.91[0.26: 1.56]	6.4×10 ⁻³	0.01[-0.56: 0.57]	9.8×10 ⁻¹	0.25[-0.12: 0.62]	1.8×10 ⁻¹
M-VLDL-CE-%	Cholesterol esters to total lipids ratio in medium VLDL	-0.03[-2.11: 2.05]	9.8×10 ⁻¹	0.69[-2.21: 3.58]	6.4×10 ⁻¹	1.74[-0.87: 4.35]	1.9×10 ⁻¹	0.66[-0.75: 2.08]	3.6×10 ⁻¹	0.28[-0.46: 1.03]	4.5×10 ⁻¹	1.14[0.49: 1.80]	6.3×10 ⁻⁴	0.24[-0.33: 0.82]	4.1×10 ⁻¹	0.55[0.17: 0.92]	4.1×10 ⁻³
M-VLDL-FC-%	Free cholesterol to total lipids ratio in medium VLDL	-0.64[-2.74: 1.46]	5.5×10 ⁻¹	2.10[-0.70: 4.91]	1.4×10 ⁻¹	-2.51[-5.08: 0.06]	5.7×10 ⁻²	-0.51[-1.92: 0.90]	4.8×10 ⁻¹	-0.58[-1.32: 0.16]	1.2×10 ⁻¹	-1.52[-2.13:-0.91]	1.1×10 ⁻⁶	-0.96[-1.50:-0.42]	5.2×10 ⁻⁴	-1.06[-1.41:-0.71]	4.6×10 ⁻⁹
M-VLDL-TG-%	Triglycerides to total lipids ratio in medium VLDL	1.07[-1.01: 3.15]	3.1×10 ⁻¹	-1.25[-4.16: 1.66]	4.0×10 ⁻¹	-2.15[-4.78: 0.48]	1.1×10 ⁻¹	-0.43[-1.85: 0.99]	5.5×10 ⁻¹	0.08[-0.68: 0.85]	8.3×10 ⁻¹	-0.87[-1.52:-0.23]	8.3×10 ⁻³	-0.05[-0.61: 0.52]	8.7×10 ⁻¹	-0.29[-0.66: 0.08]	1.3×10 ⁻¹
S-VLDL-PL-%	Phospholipids to total lipids ratio in small VLDL	0.53[-1.52: 2.58]	6.1×10 ⁻¹	-2.30[-5.14: 0.54]	1.1×10 ⁻¹	2.10[-0.46: 4.67]	1.1×10 ⁻¹	0.31[-1.08: 1.71]	6.6×10 ⁻¹	0.71[-0.02: 1.44]	5.8×10 ⁻²	0.87[0.23: 1.50]	7.4×10 ⁻³	-0.05[-0.61: 0.51]	8.7×10 ⁻¹	0.44[0.08: 0.81]	1.8×10 ⁻²
S-VLDL-C-%	Total cholesterol to total lipids ratio in small VLDL	1.21[-0.86: 3.27]	2.5×10 ⁻¹	0.23[-2.54: 3.01]	8.7×10 ⁻¹	2.03[-0.55: 4.61]	1.2×10 ⁻¹	1.20[-0.19: 2.59]	9.1×10 ⁻²	0.17[-0.57: 0.92]	6.5×10 ⁻¹	0.85[0.21: 1.49]	9.7×10 ⁻³	0.04[-0.51: 0.59]	8.8×10 ⁻¹	0.33[-0.03: 0.70]	7.2×10 ⁻²
S-VLDL-CE-%	Cholesterol esters to total lipids ratio in small VLDL	1.57[-0.49: 3.62]	1.4×10 ⁻¹	0.54[-2.20: 3.28]	7.0×10 ⁻¹	1.96[-0.60: 4.53]	1.4×10 ⁻¹	1.42[0.04: 2.80]	4.4×10 ⁻²	0.20[-0.53: 0.93]	5.9×10 ⁻¹	0.87[0.24: 1.50]	7.0×10 ⁻³	0.16[-0.38: 0.71]	5.6×10 ⁻¹	0.40[0.04: 0.76]	2.9×10 ⁻²
S-VLDL-FC-%	Free cholesterol to total lipids ratio in small VLDL	-1.87[-3.95: 0.21]	7.9×10 ⁻²	0.60[-2.33: 3.52]	6.9×10 ⁻¹	0.54[-2.09: 3.17]	6.9×10 ⁻¹	-0.58[-2.00: 0.85]	4.3×10 ⁻¹	-0.12[-0.89: 0.66]	7.7×10 ⁻¹	-0.04[-0.69: 0.61]	9.1×10 ⁻¹	-0.90[-1.44:-0.37]	9.5×10 ⁻⁴	-0.46[-0.82:-0.09]	1.4×10 ⁻²
S-VLDL-TG-%	Triglycerides to total lipids ratio in small VLDL	-1.12[-3.19: 0.94]	2.9×10 ⁻¹	-0.55[-3.44: 2.34]	7.1×10 ⁻¹	-2.67[-5.27:-0.08]	4.5×10 ⁻²	-1.44[-2.85:-0.03]	4.5×10 ⁻²	-0.46[-1.23: 0.31]	2.4×10 ⁻¹	-1.14[-1.79:-0.50]	5.6×10 ⁻⁴	-0.10[-0.66: 0.46]	7.3×10 ⁻¹	-0.52[-0.90:-0.15]	5.6×10 ⁻³
XS-VLDL-PL-%	Phospholipids to total lipids ratio in very small VLDL	4.40[2.37: 6.42]	2.4×10 ⁻⁵	1.25[-1.58: 4.09]	3.9×10 ⁻¹	2.19[-0.42: 4.81]	1.0×10 ⁻¹	3.01[1.61: 4.40]	2.3×10 ⁻⁵	1.33[0.55: 2.12]	8.5×10 ⁻⁴	0.96[0.34: 1.59]	2.5×10 ⁻³	0.73[0.18: 1.28]	9.3×10 ⁻³	0.94[0.58: 1.31]	4.3×10 ⁻⁷
XS-VLDL-C-%	Total cholesterol to total lipids ratio in very small VLDL	5.30[3.28: 7.31]	3.3×10 ⁻⁷	-0.27[-3.15: 2.61]	8.6×10 ⁻¹	2.07[-0.56: 4.70]	1.2×10 ⁻¹	3.08[1.68: 4.47]	1.6×10 ⁻⁵	0.69[-0.07: 1.45]	7.6×10 ⁻²	0.93[0.30: 1.55]	3.6×10 ⁻³	0.51[-0.03: 1.06]	6.6×10 ⁻²	0.69[0.33: 1.05]	1.7×10 ⁻⁴
XS-VLDL-CE-%	Cholesterol esters to total lipids ratio in very small VLDL	6.15[4.16: 8.15]	2.5×10 ⁻⁹	0.01[-2.90: 2.91]	1.0×10 ⁺⁰	2.38[-0.21: 4.98]	7.4×10 ⁻²	3.67[2.28: 5.06]	2.3×10 ⁻⁷	0.71[-0.04: 1.45]	6.2×10 ⁻²	0.91[0.30: 1.52]	3.7×10 ⁻³	0.67[0.12: 1.21]	1.7×10 ⁻²	0.76[0.40: 1.12]	3.1×10 ⁻⁵
XS-VLDL-FC-%	Free cholesterol to total lipids ratio in very small VLDL	-1.61[-3.68: 0.45]	1.3×10 ⁻¹	-0.79[-3.53: 1.95]	5.7×10 ⁻¹	-0.68[-3.33: 1.97]	6.2×10 ⁻¹	-1.14[-2.54: 0.26]	1.1×10 ⁻¹	0.02[-0.77: 0.81]	9.6×10 ⁻¹	0.03[-0.61: 0.66]	9.4×10 ⁻¹	-0.85[-1.38:-0.31]	1.9×10 ⁻³	-0.38[-0.74:-0.02]	4.0×10 ⁻²
XS-VLDL-TG-%	Triglycerides to total lipids ratio in very small VLDL	-5.83[-7.83: -3.83]	1.7×10 ⁻⁸	-0.03[-2.91: 2.85]	9.8×10 ⁻¹	-3.04[-5.62:-0.46]	2.2×10 ⁻¹	-3.68[-5.07:-2.30]	1.9×10 ⁻⁷	-0.94[-1.72:-0.16]	1.9×10 ⁻²	-1.50[-2.14:-0.85]	5.2×10 ⁻⁶	-0.76[-1.32:-0.20]	7.9×10 ⁻¹	-1.05[-1.42:-0.68]	3.3×10 ⁻⁸
IDL-PL-%	Phospholipids to total lipids ratio in IDL	-2.10[-4.19: -0.02]	4.8×10 ⁻²	-2.16[-4.82: 0.50]	1.1×10 ⁻¹	1.48[-1.14: 4.10]	2.7×10 ⁻¹	-1.11[-2.50: 0.28]	1.2×10 ⁻¹	0.59[-0.15: 1.32]	1.2×10 ⁻¹	0.68[0.06: 1.30]	3.1×10 ⁻²	0.91[0.34: 1.48]	1.7×10 ⁻³	0.75[0.39: 1.12]	5.0×10 ⁻⁵
IDL-C-%	Total cholesterol to total lipids ratio in IDL	8.96[7.04: 10.09]	9.1×10 ⁻¹⁹	1.23[-1.52: 3.98]	3.8×10 ⁻¹	2.78[0.21: 5.35]	3.6×10 ⁻²	5.43[4.08: 6.77]	2.6×10 ⁻¹⁵	0.46[-0.29: 1.21]	2.3×10 ⁻¹	1.56[0.96: 2.17]	4.9×10 ⁻⁷	0.46[-0.06: 0.99]	8.5×10 ⁻²	0.83[0.48: 1.18]	3.6×10 ⁻⁶
IDL-CE-%	Cholesterol esters to total lipids ratio in IDL	9.23[7.31: 11.15]	7.6×10 ⁻²⁰	1.60[-1.16: 4.36]	2.6×10 ⁻¹	1.97[-0.61: 4.55]	1.4×10 ⁻¹	5.45[4.10: 6.79]	2.1×10 ⁻¹⁵	0.40[-0.34: 1.13]	2.9×10 ⁻¹	1.18[0.58: 1.78]	1.3×10 ⁻⁴	0.43[-0.11: 0.96]	1.2×10 ⁻¹	0.67[0.32: 1.03]	1.7×10 ⁻⁴
IDL-FC-%	Free cholesterol to total lipids ratio in IDL	3.12[1.08: 5.16]	2.8×10 ⁻³	1.08[-1.84: 4.00]	4.7×10 ⁻¹	1.76[-0.90: 4.43]	2.0×10 ⁻¹	2.26[0.84: 3.67]	1.8×10 ⁻³	0.70[-0.11: 1.50]	8.9×10 ⁻²	0.70[0.05: 1.34]	3.4×10 ⁻²	0.20[-0.35: 0.76]	4.8×10 ⁻¹	0.48[0.10: 0.85]	1.2×10 ⁻²
IDL-TG-%	Triglycerides to total lipids ratio in IDL	-9.26[-11.18: -7.34]	6.2×10 ⁻²⁰	-1.31[-4.10: 1.48]	3.6×10 ⁻¹	-3.16[-5.76:-0.55]	1.9×10 ⁻²	-5.75[-7.10:-4.40]	8.2×10 ⁻¹⁷	-0.89[-1.66:-0.13]	2.2×10 ⁻²	-1.84[-2.46:-1.22]	6.6×10 ⁻⁹	-0.77[-1.30:-0.24]	4.7×10 ⁻³	-1.15[-1.51:-0.80]	2.4×10 ⁻¹⁰
L-LDL-PL-%	Phospholipids to total lipids ratio in large LDL	-3.58[-5.61: -1.55]	5.9×10 ⁻⁴	-1.52[-4.18: 1.13]	2.6×10 ⁻¹	0.66[-1.98: 3.30]	6.2×10 ⁻¹	-1.87[-3.25:-0.50]	7.7×10 ⁻³	0.15[-0.58: 0.89]	6.8×10 ⁻¹	0.34[-0.26: 0.94]	2.7×10 ⁻¹	0.37[-0.16: 0.91]	1.7×10 ⁻¹	0.31[-0.04: 0.66]	8.2×10 ⁻²
L-LDL-C-%	Total cholesterol to total lipids ratio in large LDL	7.73[5.77: 9.69]	4.0×10 ⁻¹⁴	1.95[-0.76: 4.65]	1.6×10 ⁻¹	2.23[-0.35: 4.80]	9.2×10 ⁻¹	4.78[3.42: 6.13]	4.3×10 ⁻¹²	0.13[-0.60: 0.86]	7.3×10 ⁻¹	0.97[0.37: 1.58]	1.6×10 ⁻³	0.33[-0.20: 0.86]	2.2×10 ⁻¹	0.50[0.15: 0.85]	4.8×10 ⁻³
L-LDL-CE-%	Cholesterol esters to total lipids ratio in large LDL	7.11[5.14: 9.09]	4.2×10 ⁻¹²	0.98[-1.60: 3.57]	1.1×10 ⁻¹	0.98[-1.60: 3.57]	4.6×10 ⁻¹	4.18[2.83: 5.54]	1.5×10 ⁻⁹	0.15[-0.58: 0.88]	6.8×10 ⁻¹	0.49[-0.12: 1.09]	1.1×10 ⁻¹	0.14[-0.42: 0.69]	6.2×10 ⁻¹	0.26[-0.09: 0.62]	1.5×10 ⁻¹
L-LDL-FC-%	Free cholesterol to total lipids ratio in large LDL	1.54[-0.54: 3.62]	1.5×10 ⁻¹	-1.74[-4.54: 1.05]	2.2×10 ⁻¹	2.53[-0.08: 5.14]	5.9×10 ⁻²	1.00[-0.41: 2.40]	1.6×10 ⁻¹	0.79[0.04: 1.55]	3.9×10 ⁻²	0.89[0.25: 1.52]	6.4×10 ⁻³	0.38[-0.21: 0.97]	2.1×10 ⁻¹	0.66[0.28: 1.04]	5.8×10 ⁻⁴
L-LDL-TG-%	Triglycerides to total lipids ratio in large LDL	-9.26[-11.19: -7.33]	8.9×10 ⁻²⁰	-1.62[-4.34: 1.10]	2.4×10 ⁻¹	-3.01[-5.59:-0.43]	2.3×10 ⁻²	-5.70[-7.04:-4.36]	9.1×10 ⁻¹⁷	-0.69[-1.44: 0.06]	7.3×10 ⁻²	-1.58[-2.19:-0.98]	3.6×10 ⁻⁷	-0.63[-1.16:-0.11]	1.9×10 ⁻²	-0.96[-1.32:-0.61]	7.7×10 ⁻⁸
M-LDL-PL-%	Phospholipids to total lipids ratio in medium LDL	-1.69[-3.73: 0.35]	1.1×10 ⁻¹	-2.07[-4.76: 0.63]	1.3×10 ⁻¹	0.07[-2.60: 2.74]	9.6×10 ⁻¹	-1.31[-2.70: 0.08]	6.4×10 ⁻¹	-0.11[-0.85: 0.63]	7.7×10 ⁻¹	0.22[-0.38: 0.83]	4.7×10 ⁻¹	0.46[-0.07: 0.98]	8.8×10 ⁻²	0.25[-0.10: 0.60]	1.5×10 ⁻¹
M-LDL-C-%	Total cholesterol to total lipids ratio in medium LDL	5.09[3.09: 7.10]	8.0×10 ⁻⁷	2.74[-0.08: 5.56]	5.7×10 ⁻²	1.50[-1.13: 4.13]	2.7×10 ⁻¹	3.53[2.14: 4.91]	6.3×10 ⁻⁷	-0.09[-0.82: 0.65]	8.2×10 ⁻¹	0.48[-0.12: 1.08]	1.2×10 ⁻¹	0.05[-0.48: 0.58]	8.6×10 ⁻¹	0.16[-0.19: 0.51]	3.6×10 ⁻¹
M-LDL-CE-%	Cholesterol esters to total lipids ratio in medium LDL	3.49[1.48: 5.51]	7.2×10 ⁻⁴	3.09[0.28: 5.91]	3.1×10 ⁻²	1.08[-1.56: 3.72]	4.3×10 ⁻¹	2.72[1.33: 4.12]	1.3×10 ⁻⁴	-0.08[-0.81: 0.65]	8.3×10 ⁻¹	0.22[-0.38: 0.82]	4.8×10 ⁻¹	-0.13[-0.66: 0.40]	6.3×10 ⁻¹	-0.00[-0.35: 0.35]	1.0×10 ⁺⁰
M-LDL-FC-%	Free cholesterol to total lipids ratio in medium LDL	-0.77[-2.82: 1.28]	4.6×10 ⁻¹	-2.61[-5.30: 0.07]	5.6×10 ⁻²	0.65[-2.02: 3.32]	6.4×10 ⁻¹	-0.88[-2.27: 0.51]	2.2×10 ⁻¹	0.05[-0.69: 0.79]	8.9×10 ⁻¹	0.75[0.14: 1.36]	1.6×10 ⁻²	0.63[-0.07: 1.19]	2.7×10 ⁻²	0.53[-0.18: 0.89]	3.5×10 ⁻³
M-LDL-TG-%	Triglycerides to total lipids ratio in medium LDL	-8.87[-10.81: -6.93]	3.7×10 ⁻¹⁸	-1.36[-4.15: 1.43]	3.4×10 ⁻¹	-2.85[-5.43:-0.26]	3.2×10 ⁻²	-5.43[-6.79:-4.08]	4.1×10 ⁻¹⁵	-0.57[-1.32: 0.18]	1.4×10 ⁻¹	-1.42[-2.03:-0.82]	4.5×10 ⁻⁶	-0.43[-0.96: 0.10]	1.1×10 ⁻¹	-0.80[-1.15:-0.44]	9.3×10 ⁻⁶
S-LDL-PL-%	Phospholipids to total lipids ratio in small LDL	-1.82[-3.86: 0.22]	8.1×10 ⁻²	-2.62[-5.27: 0.03]	5.3×10 ⁻²	-0.45[-3.11: 2.22]	7.4×10 ⁻¹	-1.67[-3.05:-0.29]	1.8×10 ⁻²	0.22[-0.51: 0.95]	5.5×10 ⁻¹	0.33[-0.27: 0.93]	2.8×10 ⁻¹	0.66[0.13: 1.19]	1.5×10 ⁻²	0.45[0.10: 0.80]	1.2×10 ⁻²
S-LDL-C-%	Total cholesterol to total lipids ratio in small LDL	5.18[3.17: 7.18]	5.5×10 ⁻⁷	2.80[-0.01: 5.61]	5.0×10 ⁻²	1.44[-1.20: 4.07]	2.9×10 ⁻¹	3.56[2.17: 4.95]	4.9×10 ⁻⁷	0.06[-0.67: 0.79]	8.8×10 ⁻¹	0.45[-0.16: 1.05]	1.5×10 ⁻¹	-0.14[-0.67: 0.38]	5.9×10 ⁻¹	0.10[-0.25: 0.45]	5.8×10 ⁻¹
S-LDL-CE-%	Cholesterol esters to total lipids ratio in small LDL	3.44[1.42: 5.46]	9.0×10 ⁻⁴	3.27[-1.48: 6.06]	2.2×10 ⁻²	1.37[-1.28: 4.02]	3.1×10 ⁻¹	2.83[1.43: 4.22]	6.9×10 ⁻⁵	-0.06[-0.79: 0.66]	7.7×10 ⁻¹	0.26[-0.34: 0.87]	3.9×10 ⁻¹	-0.22[-0.74: 0.31]	4.1×10 ⁻¹	0.02[-0.37: 0.33]	9.0×10 ⁻¹
S-LDL-FC-%	Free cholesterol to total lipids ratio in small LDL	-1.28[-3.33: 0.78]	2.2×10 ⁻¹	-3.31[-5.96:-0.65]	1.5×10 ⁻²	-0.04[-2.74: 2.65]	9.7×10 ⁻¹	-1.50[-2.90:-0.11]	3.4×10 ⁻²	0.02[-0.71: 0.75]	9.5×10 ⁻¹	0.50[-0.11: 1.11]	1.1×10 ⁻¹	0.40[-0.16: 0.95]	1.6×10 ⁻¹	0.34[-0.02: 0.70]	6.1×10 ⁻²
S-LDL-TG-%	Triglycerides to total lipids ratio in small LDL	-8.13[-10.10: -6.17]	2.5×10 ⁻¹⁵	-0.78[-3.62: 2.06]	5.9×10 ⁻¹	-3.06[-5.66:-0.46]	2.2×10 ⁻²	-5.00[-6.37:-3.63]	8.9×10 ⁻¹³	-0.51[-1.27: 0.24]	1.8×10 ⁻¹	-1.67[-2.28:-1.05]	1.5×10 ⁻⁷	-0.58[-1.11:-0.04]	3.4×10 ⁻²	-0.92[-1.28:-0.57]	3.7×10 ⁻⁷
XL-HDL-PL-%	Phospholipids to total lipids ratio in very large HDL	-2.14[-4.31: 0.03]	5.3×10 ⁻²	2.44[-5.06: 1.09]	9.6×10 ⁻¹	0.07[-2.77: 2.90]	9.6×10 ⁻¹	-0.16[-1.60: 1.27]	1.2×10 ⁻²	0.93[0.20: 1.65]	1.2×10 ⁻²	1.06[0.41: 1.70]	1.3×10 ⁻³	0.23[-0.35: 0.81]	4.3×10 ⁻¹	0.69[0.32: 1.06]	2.8×10 ⁻⁴
XL-HDL-C-%	Total cholesterol to total lipids ratio in very large HDL	3.19[1.05: 5.34]	3.6×10 ⁻³	-1.71[-4.70: 1.29]	2.7×10 ⁻¹	-1.87[-4.44: 0.70]	1.6×10 ⁻¹	0.46[-0.98: 1.91]	5.3×10 ⁻¹	-0.80[-1.52:-0.08]	3.0×10 ⁻²	-0.80[-1.44:-0.15]	1.5×10 ⁻²	-0.56[-1.12: 0.00]	5.0×10 ⁻²	-0.70[-1.06:-0.33]	1.8×10 ⁻⁴
XL-HDL-CE-%	Cholesterol esters to total lipids ratio in very large HDL	2.30[0.13: 4.46]	3.8×10 ⁻²	-2.03[-5.05: 0.99]	1.9×10 ⁻¹	-1.72[-4.29: 0.86]	1.9×10 ⁻¹	0.02[-1.44: 1.47]	9.8×10 ⁻¹	-0.82[-1.55:-0.08]	2.9×10 ⁻²	-1.07[-1.72:-0.42]	1.2×10 ⁻³	-0.54[-1.10: 0.02]	6.1×10 ⁻²	-0.78[-1.15:-0.41]	3.3×10 ⁻⁵
XL-HDL-FC-%	Free cholesterol to total lipids ratio in very large HDL	3.07[0.93: 5.21]	5.0×10 ⁻³	-0.30[-3.19: 2.58]	8.4×10 ⁻¹	-0.47[-3.08: 2.15]	7.3×10 ⁻¹	1.17[-0.27: 2.60]									

Fw6/FA	Ratio of omega-6 fatty acids to total fatty acids	2.66[0.61: 4.71]	1.1×10 ⁻²	-0.06[-2.93: 2.80]	9.6×10 ⁻¹	2.11[-0.52: 4.74]	1.2×10 ⁻¹	1.84[0.44: 3.25]	1.0×10 ⁻²	0.52[-0.25: 1.28]	1.9×10 ⁻¹	1.23[0.60: 1.86]	1.5×10 ⁻⁴	-0.19[-0.74: 0.37]	5.1×10 ⁻¹	0.45[0.08: 0.82]	1.6×10 ⁻²
PUFA/FA	Ratio of polyunsaturated fatty acids to total fatty acids	2.64[0.59: 4.70]	1.2×10 ⁻²	0.54[-2.36: 3.43]	7.2×10 ⁻¹	1.94[-0.69: 4.57]	1.5×10 ⁻¹	1.94[0.53: 3.35]	7.2×10 ⁻³	1.15[0.37: 1.92]	3.8×10 ⁻³	1.21[0.59: 1.84]	1.5×10 ⁻⁴	-0.33[-0.88: 0.22]	2.4×10 ⁻¹	0.52[0.16: 0.89]	5.1×10 ⁻³
MUFA/FA	Ratio of monounsaturated fatty acids to total fatty acids	-5.72[-7.73: -3.70]	4.0×10 ⁻⁸	0.11[-2.86: 3.07]	9.4×10 ⁻¹	-3.30[-5.95:-0.64]	1.6×10 ⁻²	-3.71[-5.12:-2.30]	2.6×10 ⁻⁷	-1.40[-2.19:-0.60]	5.8×10 ⁻⁴	-1.83[-2.46:-1.19]	2.3×10 ⁻⁸	-0.66[-1.24:-0.08]	2.5×10 ⁻²	-1.23[-1.61:-0.86]	1.4×10 ⁻¹⁰
SFA/FA	Ratio of saturated fatty acids to total fatty acids	4.57[2.55: 6.59]	1.1×10 ⁻⁵	-0.89[-3.66: 1.88]	5.3×10 ⁻¹	1.43[-1.22: 4.08]	2.9×10 ⁻¹	2.33[0.93: 3.72]	1.0×10 ⁻³	0.16[-0.57: 0.88]	6.7×10 ⁻¹	0.52[-0.09: 1.12]	9.6×10 ⁻²	1.05[0.53: 1.57]	7.2×10 ⁻⁵	0.67[0.33: 1.02]	1.4×10 ⁻⁴
VLDL-D	Mean diameter for VLDL particles	0.72[-1.35: 2.79]	5.0×10 ⁻¹	1.91[-1.18: 4.99]	2.3×10 ⁻¹	-2.06[-4.64: 0.52]	1.2×10 ⁻¹	0.12[-1.31: 1.55]	8.7×10 ⁻¹	-0.83[-1.63:-0.03]	4.3×10 ⁻²	-0.84[-1.49:-0.18]	1.3×10 ⁻²	-0.15[-0.75: 0.44]	6.1×10 ⁻¹	-0.55[-0.93:-0.16]	5.6×10 ⁻³
LDL-D	Mean diameter for LDL particles	-2.58[-4.66: -0.50]	1.5×10 ⁻²	-2.66[-5.44: 0.11]	6.0×10 ⁻²	0.36[-2.31: 3.02]	7.9×10 ⁻¹	-1.78[-3.19:-0.37]	1.4×10 ⁻²	0.00[-0.73: 0.73]	9.9×10 ⁻¹	-0.94[-1.54:-0.34]	2.2×10 ⁻³	-0.13[-0.68: 0.42]	6.4×10 ⁻¹	-0.38[-0.73:-0.03]	3.5×10 ⁻²
HDL-D	Mean diameter for HDL particles	2.02[-0.12: 4.15]	6.4×10 ⁻²	0.82[-2.47: 4.11]	6.2×10 ⁻¹	3.67[1.03: 6.31]	7.2×10 ⁻³	2.29[0.81: 3.78]	2.4×10 ⁻³	1.38[0.52: 2.24]	1.6×10 ⁻³	1.90[1.20: 2.61]	1.2×10 ⁻⁷	0.56[-0.10: 1.23]	9.7×10 ⁻²	1.24[0.82: 1.66]	7.5×10 ⁻⁹

APPENDIX D

The faecal metabolome as a functional readout of the gut microbiome

Figure D.1 The effect of storage time on faecal metabolites. To assess the effect of storage (a) in the participants' fridge before being stored in our biobank and (b) in the freezer at -80°C before being analysed, I calculated regression models of faecal metabolites against both storage times. Here I present qq-plots where the dashed lines indicate Bonferroni cut-off.

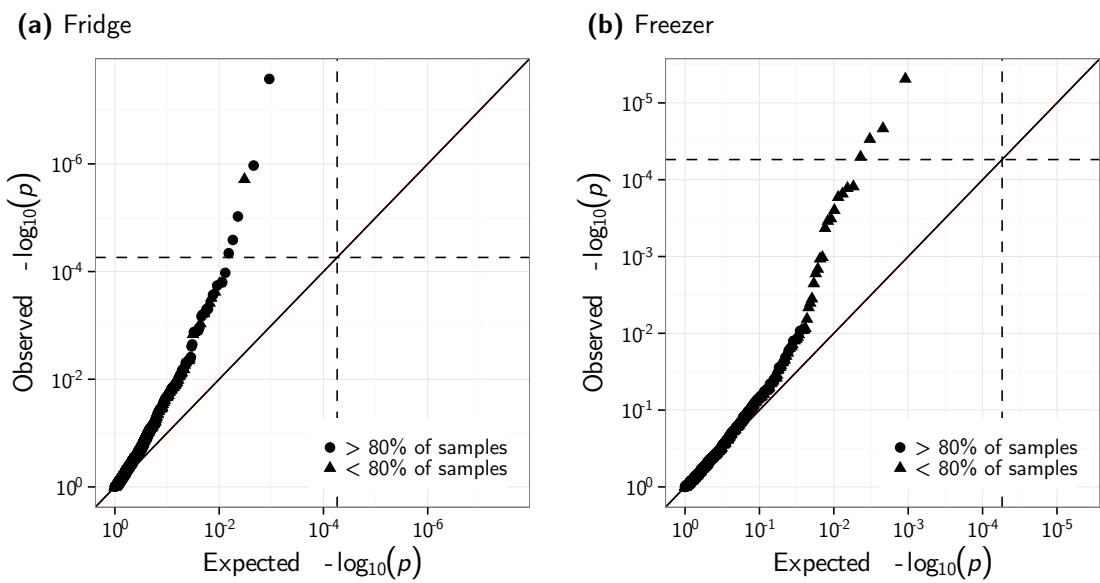


Figure D.2 Variance components of the faecal metabolome. I estimated heritability (red) and the effect of common environment (blue) using ACE models from 148 pairs of MZ twins and 155 pairs of DZ twins. Additionally, I estimated the proportion of variance explained by the gut microbiome (green) using mixed models incorporating the microbial beta-diversity for 644 individuals with 16S sequencing data available. The heat-map panels show associations of faecal metabolites with microbial alpha diversity ($n = 644$, red), visceral fat mass ($n = 647$, blue) and BMI ($n = 786$, green), where darker colours indicate stronger associations and grey indicates non-significant associations (FDR corrected). This plot illustrates results for 345 metabolites observed in less than 80 % but more than 20 % of the samples. Results for more common metabolites are shown in Figure 7.4

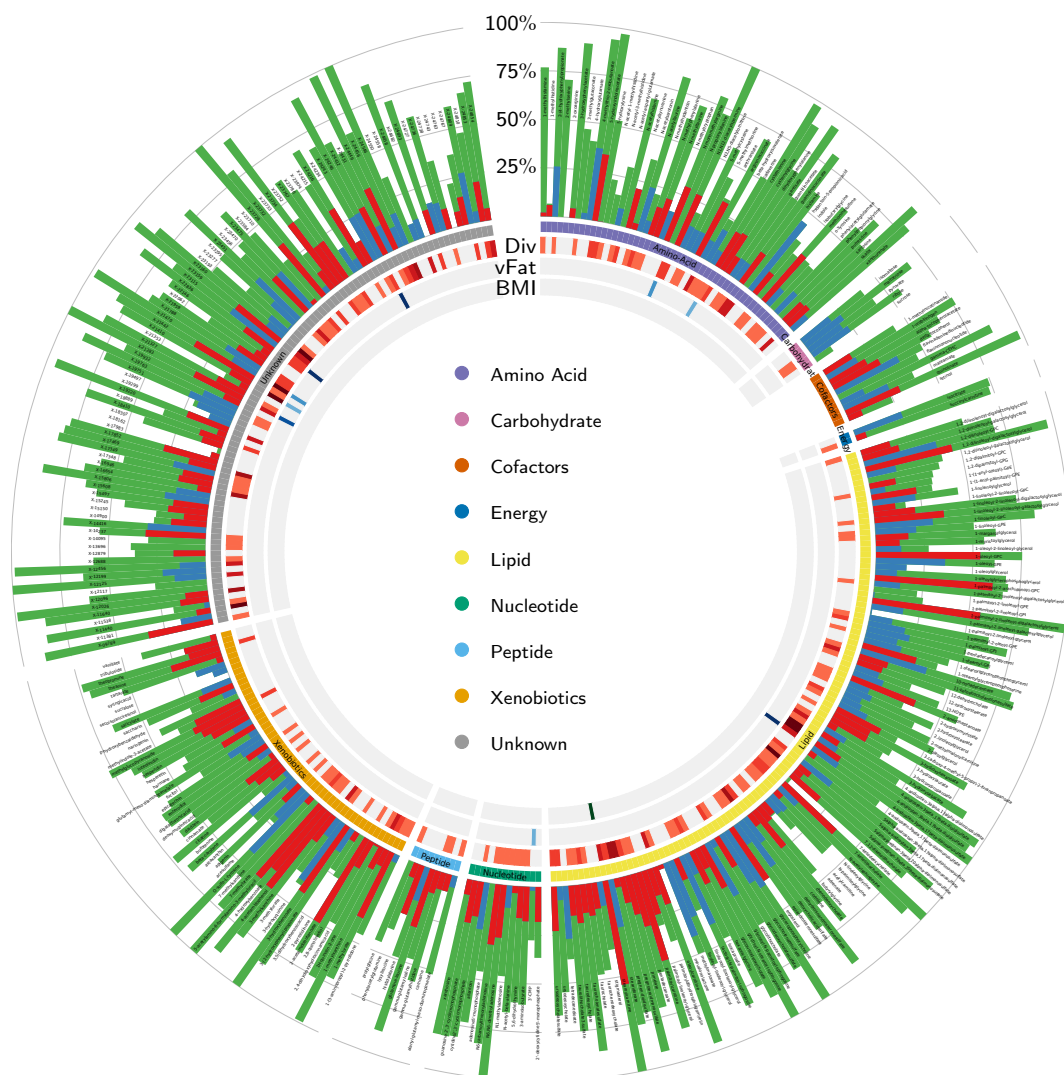
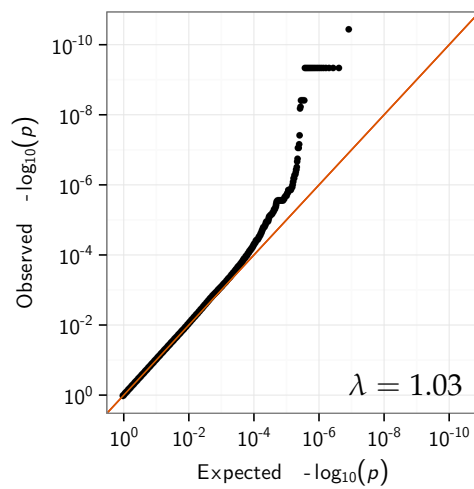
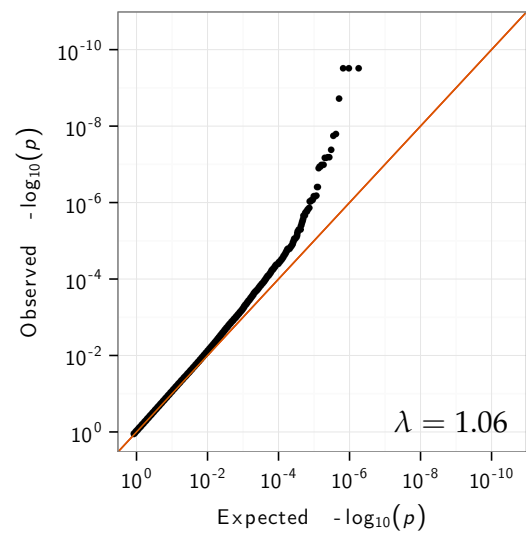


Figure D.3 Faecal metabolic traits associated with host genetics. Each panel shows the qq-plot for one of the three metabolites ((a), (b), (c)) and the metabolite ratio ((d)) with genome-wide significant associations.

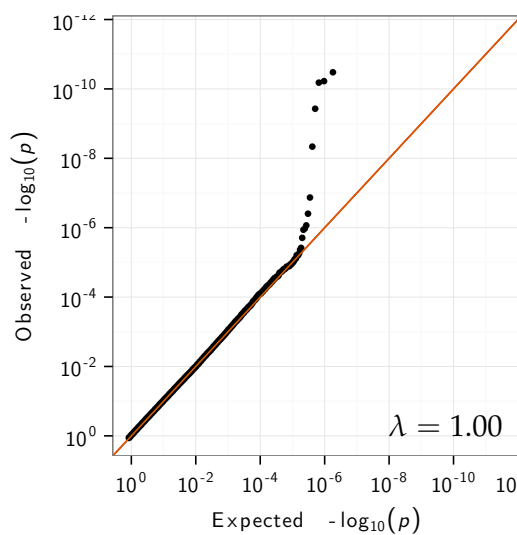
(a) 3-phenylpropionate



(b) Eicosapentaenoate



(c) 3-hydroxyhexanoate



(d) 1,3-dimethylurate / 5-acetylamino-6-amino-3-methyluracil

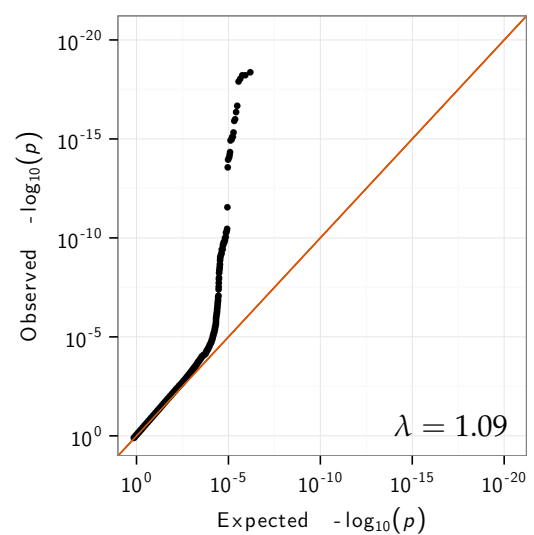


Figure D.4 Genetic loci associated with faecal metabolic traits. I found three faecal metabolites and one metabolite ratio significantly associated with genetic loci. Each panel shows one associations of a metabolic trait with the lead SNP. 3-hydroxyhexanoate was found in less than 80 % of all samples and was, thus, analysed as dichotomous trait. The other metabolites are observed in at least 80 % of the samples and were analysed continuously.

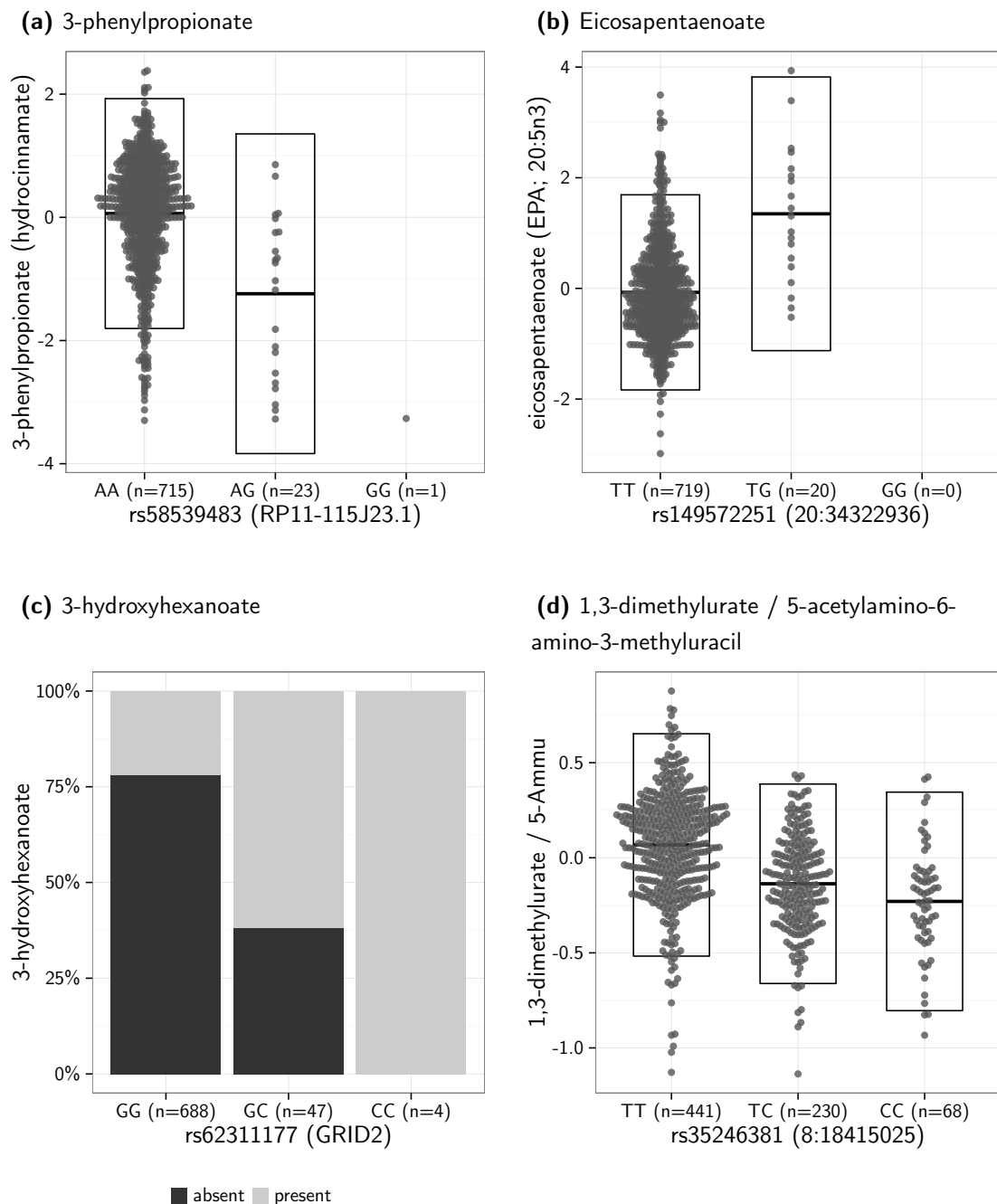
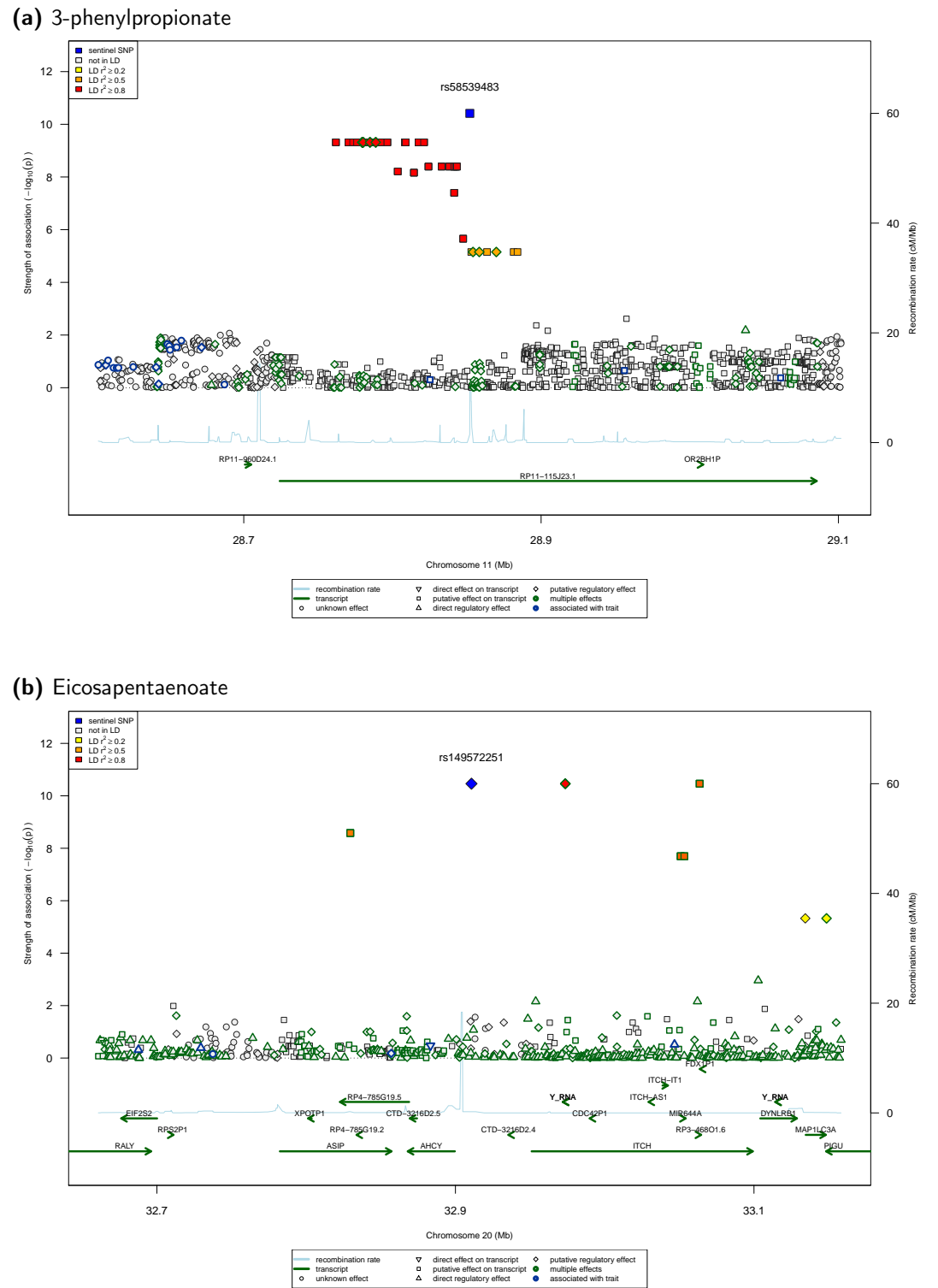
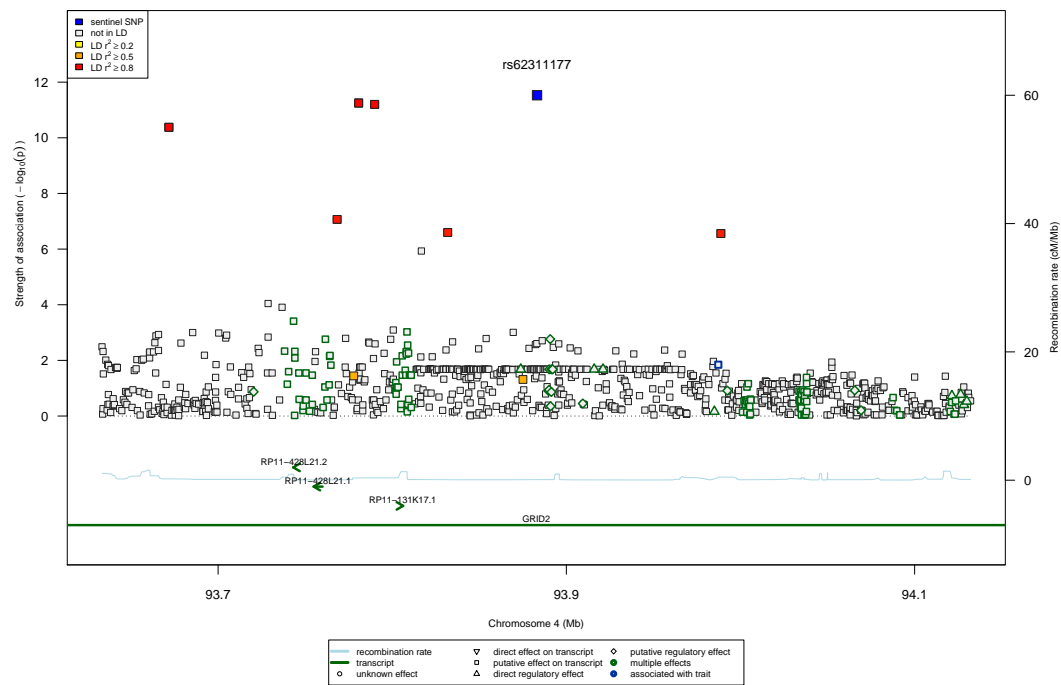


Figure D.5 Regional association plot of faecal metabolites. Regional association plots were created for all significant associations of genetic loci with faecal metabolites using the web tool SNIPA (Arnold et al., 2015). Colours indicate the strength of linkage disequilibrium (LD) with the sentinel SNP. The chromosomal positions are based on GRC37 and Ensembl v82 was used for gene annotations.



(c) 3-hydroxyhexanoate



(d) 1,3-dimethylurate / 5-acetylamino-6-amino-3-methyluracil

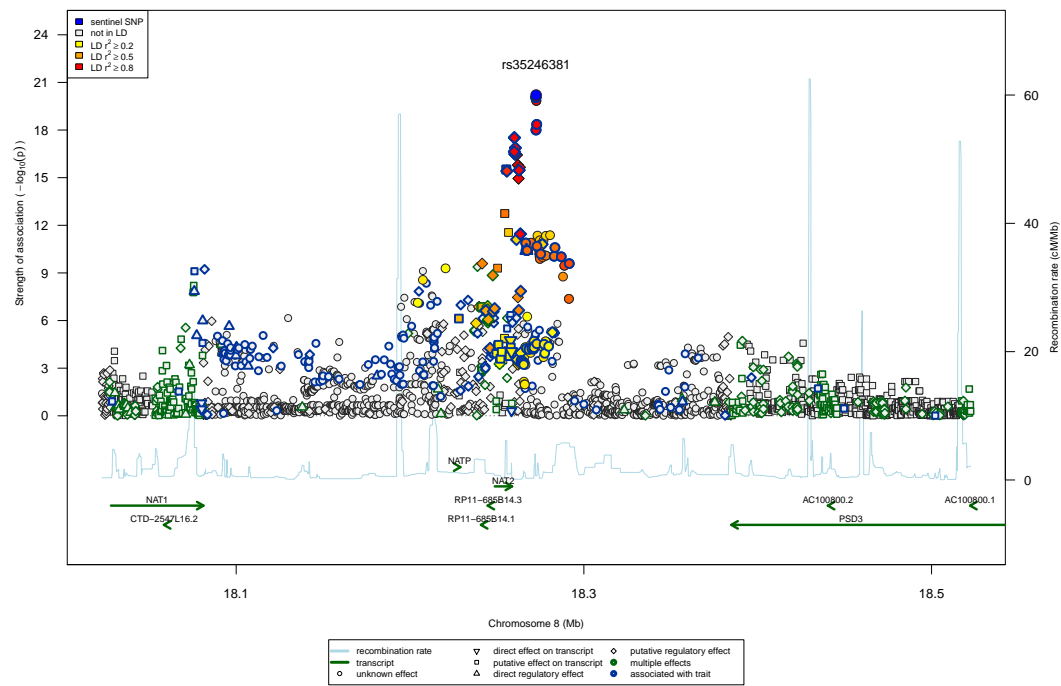


Table D.1 Variance components and phenotype associations of the faecal metabolome. In total, 915 faecal metabolites were analysed. Here, I show the proportion of samples in which the metabolite was observed (n) and its RSD for all metabolites present in at least 90 % of quality control samples. I partitioned the variance of each faecal metabolite in heritable (A), common environment (C) and unique environment (E) using structural equation modelling. Additionally, the variance explained by the microbial composition (M) was estimated using UniFrac beta diversities. This might overlap with variance components from the ACE models. To investigate associations with phenotypes – age, BMI, visceral fat mass and eGFR – I calculated linear regression models for metabolites present in more than 80 % of the samples and logistic regression models for metabolites present in less than 80 % but more than 20 % of samples.

Metabolite	N	RSD	Variance Components				Age		BMI		Visceral Fat		eGFR	
			A	C	E	M	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p	β [95 %CI]	p
X - 17162	95.3	9.4	0.24[0.10-0.38]	0.00[0.00-0.00]	0.76[0.62-0.90]	0.85[0.55-1.15]	-0.10[-0.18;-0.03]	7.3×10 ⁻³	0.14[0.08: 0.21]	1.2×10 ⁻⁵	4.19[1.57: 6.82]	1.8×10 ⁻³	0.17[-0.84: 1.17]	7.4×10 ⁻¹
X - 21410	45.8	16.0	0.12[0.00-0.34]	0.00[0.00-0.00]	0.88[0.66-1.10]	0.65[0.32-0.97]	-0.20[0.01-0.35]	9.2×10 ⁻³	0.12[0.00: 0.24]	4.2×10 ⁻²	8.40[3.54-13.26]	7.0×10 ⁻⁴	-0.94[-2.83: 0.96]	3.3×10 ⁻¹
X - 11429	93.0	5.9	0.00[0.00-0.00]	0.19[0.07-0.31]	0.81[0.69-0.93]	0.85[0.55-1.16]	-0.10[-0.18;-0.02]	1.2×10 ⁻²	0.04[-0.02: 0.11]	2.2×10 ⁻¹	5.19[2.60: 7.79]	8.9×10 ⁻⁵	0.22[-0.75: 1.20]	6.6×10 ⁻¹
X - 21470	33.6	12.6	0.00[0.00-0.00]	0.14[0.00-0.33]	0.86[0.67-1.04]	0.83[0.54-1.13]	-0.20[0.02-0.36]	1.8×10 ⁻²	0.12[-0.01: 0.24]	7.8×10 ⁻²	9.41[4.22-14.61]	3.9×10 ⁻⁴	-0.19[-2.24: 1.85]	8.5×10 ⁻¹
glycerophosphoglycerol	93.4	20.6	0.29[0.15-0.44]	0.00[0.00-0.00]	0.71[0.56-0.85]	1.00[0.72-1.27]	-0.09[-0.17;-0.01]	2.1×10 ⁻²	-0.02[-0.09: 0.04]	5.0×10 ⁻¹	4.47[1.87: 7.08]	7.6×10 ⁻⁴	0.77[-0.21: 1.76]	1.2×10 ⁻¹
glutamate, gamma-methyl ester	94.4	16.3	0.11[0.00-0.25]	0.00[0.00-0.00]	0.89[0.75-1.04]	1.00[0.76-1.24]	-0.09[-0.16;-0.01]	2.3×10 ⁻²	0.00[-0.06: 0.07]	9.1×10 ⁻¹	4.30[1.77: 6.84]	8.9×10 ⁻⁴	1.10[0.08: 2.12]	3.5×10 ⁻²
glycylvaline	100.0	8.7	0.23[0.09-0.36]	0.00[0.00-0.00]	0.77[0.64-0.91]	0.96[0.69-1.23]	-0.09[-0.16;-0.01]	2.5×10 ⁻²	0.04[-0.02: 0.10]	2.3×10 ⁻¹	4.95[2.41: 7.50]	1.4×10 ⁻⁴	0.68[-0.30: 1.65]	1.7×10 ⁻¹
mevalonate	87.8	17.6	0.19[0.04-0.33]	0.00[0.00-0.00]	0.81[0.67-0.96]	0.81[0.49-1.14]	-0.09[-0.16;-0.01]	3.5×10 ⁻²	0.06[-0.01: 0.12]	9.6×10 ⁻²	4.37[1.70: 7.04]	1.4×10 ⁻³	0.73[-0.31: 1.76]	1.7×10 ⁻¹
N-propionylalanine	49.2	31.5	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.72[0.41-1.03]	-0.15[0.04-0.29]	4.1×10 ⁻²	0.08[-0.04: 0.20]	1.7×10 ⁻¹	8.45[3.69-13.20]	5.0×10 ⁻⁴	-0.95[-2.80: 0.90]	3.1×10 ⁻¹
linoleoyl-arachidonoyl-glycerol (18:2/20:4)*	98.3	9.8	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.83[0.54-1.13]	-0.07[-0.15-0.00]	5.0×10 ⁻²	0.11[0.05: 0.17]	2.7×10 ⁻⁴	4.52[2.10: 6.93]	2.4×10 ⁻⁴	0.57[-0.37: 1.50]	2.3×10 ⁻¹
3-ketosphinganine	87.3	19.2	0.28[0.12-0.43]	0.00[0.00-0.00]	0.72[0.57-0.88]	0.75[0.41-1.10]	-0.08[-0.16: 0.00]	5.0×10 ⁻²	0.14[0.07: 0.20]	5.0×10 ⁻⁵	2.22[-0.44: 4.88]	1.0×10 ⁻¹	-0.65[-1.71: 0.41]	2.3×10 ⁻¹
beta-alanine	95.4	45.8	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	1.00[0.74-1.26]	-0.07[-0.14: 0.01]	7.0×10 ⁻²	0.02[-0.04: 0.08]	5.0×10 ⁻¹	4.29[1.88: 6.71]	5.0×10 ⁻⁴	0.30[-0.65: 1.26]	5.3×10 ⁻¹
octadecanedioate	87.4	9.4	0.33[0.19-0.48]	0.00[0.00-0.00]	0.67[0.52-0.81]	0.73[0.39-1.07]	0.08[-0.01: 0.16]	7.1×10 ⁻²	-0.01[-0.08: 0.06]	7.3×10 ⁻¹	-5.10[-7.71:-2.48]	1.3×10 ⁻⁴	-0.10[-1.12: 0.93]	8.5×10 ⁻¹
4-androsten-3beta,17beta-diol monosulfate (1)	24.6	11.1	0.00[0.00-0.00]	0.16[0.00-0.36]	0.84[0.64-1.04]	0.93[0.66-1.20]	-0.17[0.07-0.35]	7.3×10 ⁻²	0.16[0.02: 0.30]	2.2×10 ⁻²	11.72[6.00-17.43]	5.8×10 ⁻⁵	-1.04[-3.27: 1.19]	3.6×10 ⁻¹
X - 21283	41.9	—	0.28[0.05-0.50]	0.00[0.00-0.00]	0.72[0.50-0.95]	0.75[0.44-1.05]	-0.14[0.08-0.30]	8.1×10 ⁻²	0.08[-0.04: 0.20]	2.0×10 ⁻¹	9.71[4.67-14.76]	1.6×10 ⁻⁴	-0.10[-2.05: 1.84]	9.2×10 ⁻¹
X - 23637	91.2	31.9	0.15[0.00-0.31]	0.00[0.00-0.00]	0.85[0.69-1.01]	0.51[0.15-0.86]	-0.07[-0.14: 0.01]	8.8×10 ⁻²	0.01[-0.06: 0.07]	8.2×10 ⁻¹	5.04[2.42: 7.67]	1.7×10 ⁻⁴	0.77[-0.23: 1.76]	1.3×10 ⁻¹
N6,N6,N6-trimethyllysine	97.3	11.3	0.00[0.00-0.00]	0.12[0.01-0.24]	0.88[0.76-0.99]	1.00[0.74-1.26]	-0.06[-0.13: 0.01]	1.1×10 ⁻¹	0.04[-0.03: 0.10]	2.5×10 ⁻¹	6.08[3.62: 8.55]	1.3×10 ⁻⁶	0.10[-0.86: 1.06]	8.4×10 ⁻¹
X - 23652	99.0	6.5	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.84[0.55-1.13]	-0.06[-0.13: 0.01]	1.2×10 ⁻¹	0.13[0.08: 0.19]	7.1×10 ⁻⁶	0.61[-1.95: 3.16]	6.4×10 ⁻¹	-0.47[-1.42: 0.48]	3.3×10 ⁻¹
3'-CMP	26.7	—	0.00[0.00-0.00]	0.02[0.00-0.23]	0.98[0.77-1.19]	0.26[0.00-0.58]	-0.13[0.12-0.29]	1.2×10 ⁻¹	0.05[-0.08: 0.19]	4.4×10 ⁻¹	9.45[3.93-14.96]	7.9×10 ⁻⁴	0.77[-1.34: 2.88]	4.8×10 ⁻¹
thioprolin	98.7	10.5	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.68[0.36-1.00]	-0.06[-0.13: 0.02]	1.2×10 ⁻¹	0.10[0.04: 0.16]	1.4×10 ⁻³	4.86[2.37: 7.36]	1.3×10 ⁻⁴	-0.17[-1.15: 0.80]	7.3×10 ⁻¹
1-methylguanidine	100.0	6.7	0.15[0.01-0.29]	0.00[0.00-0.00]	0.85[0.71-0.99]	0.68[0.37-1.00]	0.06[-0.02: 0.13]	1.3×10 ⁻¹	0.09[0.03: 0.15]	3.6×10 ⁻³	-1.10[-3.67: 1.46]	4.0×10 ⁻¹	-2.12[-3.06:-1.19]	8.6×10 ⁻⁶
X - 24678	99.7	14.8	0.28[0.15-0.41]	0.00[0.00-0.00]	0.72[0.59-0.85]	1.00[0.75-1.25]	-0.06[-0.13: 0.02]	1.3×10 ⁻¹	0.03[-0.03: 0.10]	2.9×10 ⁻¹	7.16[4.68: 9.64]	1.6×10 ⁻⁸	-0.54[-1.50: 0.42]	2.7×10 ⁻¹
gamma-glutamyl-epsilon-lysine	99.0	8.2	0.00[0.00-0.00]	0.11[0.00-0.22]	0.89[0.78-1.00]	1.00[0.75-1.25]	-0.05[-0.13: 0.02]	1.5×10 ⁻¹	0.04[-0.02: 0.10]	1.7×10 ⁻¹	4.53[2.03: 7.04]	3.9×10 ⁻⁴	0.56[-0.39: 1.51]	2.5×10 ⁻¹
X - 23747	98.7	5.2	0.24[0.09-0.38]	0.00[0.00-0.00]	0.76[0.62-0.91]	1.00[0.76-1.24]	-0.05[-0.13: 0.02]	1.5×10 ⁻¹	0.05[-0.02: 0.11]	1.5×10 ⁻¹	5.52[3.06: 7.99]	1.1×10 ⁻⁵	0.45[-0.49: 1.39]	3.5×10 ⁻¹
N-acetylarginine	100.0	4.4	0.00[0.00-0.00]	0.17[0.06-0.28]	0.83[0.72-0.94]	1.00[0.79-1.21]	-0.05[-0.13: 0.02]	1.5×10 ⁻¹	-0.02[-0.08: 0.05]	6.0×10 ⁻¹	4.28[1.82: 6.74]	6.5×10 ⁻⁴	-0.01[-0.96: 0.94]	9.9×10 ⁻¹
histidine	100.0	7.7	0.24[0.11-0.37]	0.00[0.00-0.00]	0.76[0.63-0.89]	0.96[0.69-1.23]	-0.05[-0.13: 0.02]	1.7×10 ⁻¹	0.07[0.01: 0.13]	3.3×10 ⁻²	4.46[1.89: 7.02]	6.6×10 ⁻⁴	0.23[-0.74: 1.21]	6.4×10 ⁻¹
nicotinate	99.7	8.3	0.28[0.15-0.42]	0.00[0.00-0.00]	0.72[0.58-0.85]	1.00[0.79-1.21]	-0.05[-0.13: 0.02]	1.7×10 ⁻¹	0.04[-0.02: 0.11]	1.8×10 ⁻¹	7.38[4.91: 9.86]	5.1×10 ⁻⁹	0.62[-0.32: 1.57]	2.0×10 ⁻¹
6-hydroxynicotinate	98.0	8.5	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.93[0.66-1.21]	-0.05[-0.12: 0.02]	1.9×10 ⁻¹	0.05[-0.01: 0.11]	1.1×10 ⁻¹	4.15[1.73: 6.57]	7.9×10 ⁻⁴	0.41[-0.53: 1.35]	3.9×10 ⁻¹
gamma-glutamyl-alpha-lysine	99.5	8.7	0.24[0.10-0.37]	0.00[0.00-0.00]	0.76[0.63-0.90]	1.00[0.74-1.26]	-0.05[-0.12: 0.02]	1.9×10 ⁻¹	0.01[-0.05: 0.07]	7.2×10 ⁻¹	5.27[2.79: 7.74]	3.0×10 ⁻⁵	1.07[0.11: 2.02]	2.9×10 ⁻²
indoleacetate	99.7	6.9	0.17[0.02-0.31]	0.00[0.00-0.00]	0.83[0.69-0.98]	0.71[0.40-1.02]	0.05[-0.03: 0.12]	2.0×10 ⁻¹	0.02[-0.04: 0.09]	4.4×10 ⁻¹	4.17[1.73: 6.61]	8.2×10 ⁻⁴	0.55[-0.38: 1.47]	2.5×10 ⁻¹
mannose	96.2	22.2	0.00[0.00-0.00]	0.00[0.00-0.00]	1.00[1.00-1.00]	0.74[0.42-1.05]	-0.05[-0.12: 0.03]	2.1×10 ⁻¹	-0.04[-0.10: 0.02]	1.5×10 ⁻¹	4.97[2.57: 7.38]	5.0×10 ⁻⁵	0.49[-0.44: 1.43]	3.0×10 ⁻¹
X - 24686	99.6	11.3	0.27[0.14-0.41]	0.00[0.00-0.00]	0.73[0.59-0.86]	1.00[0.75-1.25]	-0.05[-0.12: 0.03]	2.3×10 ⁻¹	0.05[-0.02: 0.11]	1.5×10 ⁻¹	5.56[3.07: 8.05]	1.2×10 ⁻⁵	-0.36[-1.31: 0.59]	4.6×10 ⁻¹
X - 24246	61.3	7.9	0.00[0.00-0.00]	0.25[0.08-0.43]	0.75[0.57-0.92]	0.67[0.36-0.99]	-0.10[0.23-0.26]	2.3×10 ⁻¹	0.14[0.01: 0.26]	3.6×10 ⁻²	10.80[5.72-15.88]	3.1×10 ⁻⁵	0.19[-1.76: 2.14]	8.5×10 ⁻¹
pantothenate	100.0	6.5	0.32[0.19-0.45]	0.00[0.00-0.00]	0.68[0.55-0.81]	1.00[0.79-1.21]	-0.05[-0.12: 0.03]	2.4×10 ⁻¹	0.06[-0.00: 0.12]	5.7×10 ⁻²	6.12[3.60: 8.64]	1.9×10 ⁻⁶	0.67[-0.30: 1.64]	1.8×10 ⁻¹
N-acetylhistidine	100.0	4.7	0.19[0.05-0.33]	0.00[0.00-0.00]	0.81[0.67-0.95]	0.91[0.63-1.19]	-0.04[-0.12: 0.03]	2.5×10 ⁻¹	0.03[-0.03: 0.09]	3.4×10 ⁻¹	6.56[4.05: 9.08]	3.0×10 ⁻⁷	0.35[-0.61: 1.31]	4.8×10 ⁻¹
N-propionylmethionine	97.1	9.0	0.00[0.00-0.00]	0.14[0.03-0.26]	0.86[0.74-0.97]	0.87[0.57-1.16]	0.04[-0.03: 0.12]	2.5×10 ⁻¹	0.04[-0.02: 0.10]	2.4×10 ⁻¹	5.42[2.89: 7.95]	2.7×10 ⁻⁵	-0.27[-1.21: 0.68]	5.8×10 ⁻¹
N-acetylmethionine	100.0	6.3	0.00[0.00-0.00]	0.14[0.03-0.25]	0.86[0.75-0.97]	1.00[0.75-1.25]	0.04[-0.03: 0.12]	2.6×10 ⁻¹	0.06[-0.00: 0.12]	6.5×10 ⁻²	5.10[2.58: 7.62]	7.2×10 ⁻⁵	0.09[-0.88: 1.06]	8.6×10 ⁻¹
eicosapentaenoate (EPA; 20:5n3)	94.7	13.6	0.16[0.01-0.31]	0.00[0.00-0.00]	0.84[0.69-0.99]	0.67[0.33-1.01]	0.04[-0.03: 0.12]	2.6×10 ⁻¹	0.07[0.01: 0.13]	3.2×10 ⁻²	4.32[1.69: 6.95]	1.3×10 ⁻³	0.22[-0.78: 1.23]	6.7×10 ⁻¹
glutamate	100.0	4.9	0.23[0.09-0.36]	0.00[0.00-0.00]	0.77[0.64-0.91]	1.00[0.75-1.25]	-0.04[-0.12: 0.03]	2.7×10 ⁻¹	0.05[-0.01: 0.11]	1.0×10 ⁻¹	7.35[4.85: 9.85]	8.3×10 ⁻⁹	0.57[-0.41: 1.55]	2.5×10 ⁻¹
sedoheptulose	99.7	5.8	0.18[0.04-0.32]	0.00[0.00-0.00]	0.82[0.68-0.96]	0.97[0.71-1.24]	-0.04[-0.11: 0.03							

glucose	99.6	5.5	0.25[0.11:0.39]	0.00[0.00:0.00]	0.75[0.61:0.89]	1.00[0.74:1.26]	-0.04[-0.11: 0.04]	3.1×10 ⁻¹	0.01[-0.05: 0.08]	6.7×10 ⁻¹	4.88[2.37: 7.39]	1.4×10 ⁻⁴	-0.07[-1.03: 0.89]	8.9×10 ⁻¹
bilirubin (Z,Z)	91.2	17.4	0.22[0.08:0.36]	0.00[0.00:0.00]	0.78[0.64:0.92]	0.77[0.44:1.10]	-0.04[-0.12: 0.04]	3.2×10 ⁻¹	0.13[0.06: 0.19]	8.9×10 ⁻⁵	3.36[0.72: 6.00]	1.2×10 ⁻²	-0.50[-1.48: 0.48]	3.2×10 ⁻¹
D-urobilin	94.3	19.4	0.20[0.06:0.34]	0.00[0.00:0.00]	0.80[0.66:0.94]	0.81[0.50:1.12]	-0.04[-0.11: 0.04]	3.2×10 ⁻¹	0.10[-0.04: 0.16]	1.8×10 ⁻³	4.43[1.76: 7.11]	1.1×10 ⁻³	0.07[-0.90: 1.04]	8.9×10 ⁻¹
N2-acetyllysine	100.0	3.5	0.00[0.00:0.00]	0.12[0.01:0.23]	0.88[0.77:0.99]	1.00[0.75:1.25]	-0.04[-0.11: 0.04]	3.3×10 ⁻¹	0.03[-0.03: 0.09]	3.9×10 ⁻¹	5.11[2.59: 7.64]	7.3×10 ⁻⁵	0.36[-0.61: 1.33]	4.6×10 ⁻¹
uridine	100.0	9.6	0.23[0.09:0.37]	0.00[0.00:0.00]	0.77[0.63:0.91]	0.99[0.73:1.25]	-0.04[-0.11: 0.04]	3.4×10 ⁻¹	0.01[-0.05: 0.07]	7.2×10 ⁻¹	4.38[1.87: 6.88]	6.2×10 ⁻⁴	0.20[-0.77: 1.16]	6.9×10 ⁻¹
gamma-glutamylphenylalanine	97.8	5.8	0.00[0.00:0.00]	0.22[0.11:0.33]	0.78[0.67:0.89]	0.78[0.48:1.09]	0.04[0.04: 0.11]	3.6×10 ⁻¹	0.01[-0.05: 0.07]	7.3×10 ⁻¹	4.64[2.10: 7.18]	3.4×10 ⁻⁴	0.45[-0.53: 1.43]	3.7×10 ⁻¹
nicotinate ribonucleoside	99.5	8.3	0.00[0.00:0.00]	0.09[0.00:0.20]	0.91[0.80:1.02]	0.87[0.58:1.16]	-0.03[-0.11: 0.04]	3.9×10 ⁻¹	-0.03[-0.09: 0.04]	4.1×10 ⁻¹	4.17[1.70: 6.63]	9.3×10 ⁻⁴	0.93[-0.02: 1.89]	5.6×10 ⁻²
glycerate	100.0	16.1	0.26[0.12:0.40]	0.00[0.00:0.00]	0.74[0.60:0.88]	1.00[0.75:1.25]	-0.03[-0.11: 0.04]	4.1×10 ⁻¹	0.03[-0.03: 0.09]	3.6×10 ⁻¹	5.11[2.59: 7.63]	7.2×10 ⁻⁵	1.05[0.11: 2.00]	2.9×10 ⁻²
stearoylcarnitine	49.4	—	0.54[0.36:0.72]	0.00[0.00:0.00]	0.46[0.28:0.64]	0.61[0.29:0.94]	-0.07[0.43:-0.25]	4.3×10 ⁻¹	0.27[0.14: 0.40]	2.7×10 ⁻⁵	6.80[1.63:11.97]	1.0×10 ⁻²	-0.00[-2.00: 1.99]	1.0×10 ⁺⁰
lysine	100.0	5.6	0.18[0.04:0.32]	0.00[0.00:0.00]	0.82[0.68:0.96]	1.00[0.75:1.25]	-0.03[-0.10: 0.04]	4.3×10 ⁻¹	0.02[-0.04: 0.08]	4.6×10 ⁻¹	4.98[2.50: 7.46]	8.3×10 ⁻⁵	0.36[-0.59: 1.30]	4.6×10 ⁻¹
fructose	99.5	6.8	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.68[0.36:1.00]	-0.03[-0.10: 0.04]	4.3×10 ⁻¹	-0.02[-0.08: 0.04]	4.6×10 ⁻¹	4.79[2.34: 7.25]	1.3×10 ⁻⁴	0.19[-0.77: 1.15]	7.0×10 ⁻¹
formiminoglutamate	98.2	4.7	0.14[0.00:0.28]	0.00[0.00:0.00]	0.86[0.72:1.01]	0.49[0.16:0.83]	0.03[-0.04: 0.10]	4.3×10 ⁻¹	0.03[-0.03: 0.09]	3.7×10 ⁻¹	4.27[1.76: 6.78]	8.7×10 ⁻⁴	-0.16[-1.12: 0.79]	7.4×10 ⁻¹
X - 24220	99.9	5.4	0.00[0.00:0.00]	0.14[0.03:0.25]	0.86[0.75:0.97]	0.85[0.56:1.14]	-0.03[-0.10: 0.04]	4.4×10 ⁻¹	0.03[-0.03: 0.09]	3.3×10 ⁻¹	4.58[2.11: 7.05]	2.8×10 ⁻⁴	0.14[-0.82: 1.11]	7.7×10 ⁻¹
N-methylalanine	97.7	13.5	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.70[0.38:1.02]	-0.03[-0.10: 0.04]	4.4×10 ⁻¹	0.09[0.03: 0.15]	2.3×10 ⁻³	4.05[1.64: 6.47]	9.8×10 ⁻⁴	0.03[-0.89: 0.96]	9.4×10 ⁻¹
N-acetylglutamine	99.1	12.6	0.18[0.05:0.31]	0.00[0.00:0.00]	0.82[0.69:0.95]	0.68[0.36:1.00]	-0.03[-0.10: 0.05]	4.4×10 ⁻¹	0.05[-0.01: 0.11]	9.7×10 ⁻²	5.54[2.99: 8.10]	2.2×10 ⁻⁵	0.09[-0.90: 1.09]	8.5×10 ⁻¹
methionine	100.0	7.0	0.13[0.00:0.27]	0.00[0.00:0.00]	0.87[0.73:1.01]	1.00[0.75:1.25]	-0.03[-0.10: 0.04]	4.5×10 ⁻¹	0.06[0.00: 0.12]	5.0×10 ⁻²	5.57[3.09: 8.05]	1.1×10 ⁻⁵	0.19[-0.79: 1.16]	7.1×10 ⁻¹
sphingosine	100.0	10.1	0.13[0.00:0.27]	0.00[0.00:0.00]	0.87[0.73:1.01]	0.89[0.61:1.17]	-0.03[-0.10: 0.05]	4.6×10 ⁻¹	0.09[0.03: 0.15]	2.1×10 ⁻³	4.11[1.69: 6.53]	8.9×10 ⁻⁴	0.05[-0.90: 1.01]	9.1×10 ⁻¹
cystathionine	31.7	—	0.20[0.00:0.44]	0.00[0.00:0.00]	0.80[0.56:1.03]	0.77[0.47:1.08]	-0.06[0.47:-0.22]	4.7×10 ⁻¹	-0.01[-0.14: 0.12]	8.5×10 ⁻¹	8.62[3.42:13.82]	1.2×10 ⁻³	-0.19[-3.20: 0.81]	2.4×10 ⁻¹
N-alpha-acetylornithine	99.9	4.2	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.65[0.33:0.98]	-0.03[-0.10: 0.05]	4.8×10 ⁻¹	0.02[-0.04: 0.08]	5.1×10 ⁻¹	4.00[1.56: 6.43]	0.01[-0.94: 0.96]	9.8×10 ⁻¹	
aspartate	100.0	10.5	0.22[0.08:0.35]	0.00[0.00:0.00]	0.78[0.65:0.92]	0.76[0.45:1.06]	-0.03[-0.10: 0.05]	4.9×10 ⁻¹	0.10[0.04: 0.16]	2.0×10 ⁻³	5.20[2.69: 7.72]	4.9×10 ⁻⁵	0.68[-0.32: 1.67]	1.8×10 ⁻¹
riboflavin (Vitamin B2)	100.0	7.6	0.13[0.00:0.27]	0.00[0.00:0.00]	0.87[0.73:1.01]	1.00[0.77:1.23]	-0.03[-0.10: 0.05]	4.9×10 ⁻¹	0.02[-0.04: 0.08]	4.8×10 ⁻¹	4.61[2.16: 7.07]	2.3×10 ⁻⁴	0.62[-0.32: 1.57]	2.0×10 ⁻¹
proline	100.0	5.4	0.00[0.00:0.00]	0.09[0.00:0.20]	0.91[0.80:1.02]	0.67[0.35:0.99]	0.03[-0.05: 0.10]	4.9×10 ⁻¹	0.03[-0.03: 0.09]	3.9×10 ⁻¹	4.92[2.38: 7.46]	1.5×10 ⁻⁴	0.24[-0.74: 1.21]	6.3×10 ⁻¹
dihydroferulic acid	97.3	7.3	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.77[0.46:1.08]	-0.03[-0.10: 0.05]	4.9×10 ⁻¹	0.03[-0.03: 0.09]	3.7×10 ⁻¹	4.28[1.83: 6.73]	6.2×10 ⁻⁴	0.32[-0.63: 1.28]	5.0×10 ⁻¹
N-acetylalanine	100.0	9.0	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.89[0.61:1.18]	-0.03[-0.10: 0.05]	4.9×10 ⁻¹	0.03[-0.03: 0.09]	2.7×10 ⁻¹	5.45[2.92: 7.98]	2.4×10 ⁻⁵	-0.29[-1.23: 0.64]	5.4×10 ⁻¹
malate	100.0	10.3	0.21[0.07:0.34]	0.00[0.00:0.00]	0.79[0.66:0.93]	0.82[0.52:1.11]	-0.03[-0.10: 0.05]	4.9×10 ⁻¹	0.05[-0.01: 0.11]	1.0×10 ⁻¹	4.74[2.23: 7.24]	2.1×10 ⁻⁴	-0.16[-1.12: 0.80]	7.5×10 ⁻¹
threonine	100.0	7.8	0.13[0.00:0.27]	0.00[0.00:0.00]	0.87[0.73:1.01]	0.87[0.58:1.15]	-0.02[-0.10: 0.05]	5.0×10 ⁻¹	0.02[-0.04: 0.08]	4.6×10 ⁻¹	5.12[2.67: 7.57]	4.2×10 ⁻⁵	0.06[-0.89: 1.02]	8.9×10 ⁻¹
2-hydroxyglutarate	100.0	7.1	0.13[0.00:0.27]	0.00[0.00:0.00]	0.87[0.73:1.01]	0.47[0.14:0.81]	-0.02[-0.10: 0.05]	5.1×10 ⁻¹	0.03[-0.03: 0.09]	4.0×10 ⁻¹	4.39[1.87: 6.92]	6.5×10 ⁻⁴	0.88[-0.07: 1.84]	7.0×10 ⁻²
N-formylmethionine	99.1	7.4	0.20[0.07:0.34]	0.00[0.00:0.00]	0.80[0.66:0.93]	1.00[0.75:1.25]	-0.03[-0.10: 0.05]	5.1×10 ⁻¹	0.03[-0.03: 0.09]	3.8×10 ⁻¹	4.81[2.31: 7.31]	1.7×10 ⁻⁴	0.24[-0.73: 1.21]	6.3×10 ⁻¹
palmitoyl ethanolamide	100.0	8.6	0.19[0.06:0.33]	0.00[0.00:0.00]	0.81[0.67:0.94]	0.92[0.65:1.20]	-0.02[-0.10: 0.05]	5.1×10 ⁻¹	0.07[0.01: 0.13]	2.2×10 ⁻²	4.26[1.82: 6.70]	6.3×10 ⁻⁴	-0.23[-1.20: 0.73]	6.4×10 ⁻¹
1-methyladenine	99.9	6.5	0.11[0.00:0.25]	0.00[0.00:0.00]	0.89[0.75:1.03]	1.00[0.78:1.22]	-0.02[-0.10: 0.05]	5.3×10 ⁻¹	0.05[-0.01: 0.11]	1.3×10 ⁻¹	4.72[2.28: 7.16]	1.5×10 ⁻⁴	0.49[-0.46: 1.44]	3.1×10 ⁻¹
ornithine	100.0	5.6	0.22[0.08:0.36]	0.00[0.00:0.00]	0.78[0.64:0.92]	0.76[0.45:1.06]	-0.02[-0.10: 0.05]	5.4×10 ⁻¹	0.05[-0.01: 0.11]	1.1×10 ⁻¹	5.07[2.59: 7.55]	6.0×10 ⁻⁵	-0.16[-1.13: 0.81]	7.4×10 ⁻¹
xylose	91.2	12.3	0.13[0.00:0.30]	0.00[0.00:0.00]	0.87[0.70:1.03]	0.84[0.52:1.15]	-0.02[-0.10: 0.05]	5.5×10 ⁻¹	0.05[-0.01: 0.11]	1.3×10 ⁻¹	4.61[2.12: 7.09]	2.8×10 ⁻⁴	0.39[-0.55: 1.34]	4.1×10 ⁻¹
glutamine	100.0	7.6	0.17[0.03:0.31]	0.00[0.00:0.00]	0.83[0.69:0.97]	0.90[0.62:1.18]	-0.02[-0.10: 0.05]	5.6×10 ⁻¹	0.04[-0.02: 0.10]	1.9×10 ⁻¹	4.21[1.68: 6.74]	1.1×10 ⁻³	0.56[-0.43: 1.54]	2.7×10 ⁻¹
3-(4-hydroxyphenyl)lactate	100.0	5.7	0.00[0.00:0.00]	0.11[0.00:0.22]	0.89[0.78:1.01]	0.60[0.28:0.93]	-0.02[-0.09: 0.05]	5.7×10 ⁻¹	0.06[-0.00: 0.12]	5.5×10 ⁻²	4.59[2.16: 7.03]	2.2×10 ⁻⁴	0.14[-0.80: 1.08]	7.8×10 ⁻¹
N6-acetyllysine	99.2	12.0	0.00[0.00:0.00]	0.13[0.02:0.25]	0.87[0.75:0.98]	0.97[0.71:1.24]	-0.02[-0.09: 0.05]	6.0×10 ⁻¹	0.03[-0.03: 0.09]	2.9×10 ⁻¹	5.99[3.41: 8.56]	5.1×10 ⁻⁶	0.80[-0.16: 1.77]	1.0×10 ⁻¹
glycine	100.0	6.0	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.80[0.51:1.10]	0.02[-0.06: 0.09]	6.5×10 ⁻¹	0.05[-0.01: 0.11]	7.4×10 ⁻²	5.40[2.87: 7.93]	2.8×10 ⁻⁵	0.23[-0.74: 1.20]	6.4×10 ⁻¹
pseudouridine	100.0	5.5	0.15[0.01:0.29]	0.00[0.00:0.00]	0.85[0.71:0.99]	1.00[0.74:1.26]	-0.02[-0.09: 0.06]	6.6×10 ⁻¹	0.00[-0.06: 0.06]	9.8×10 ⁻¹	5.21[2.76: 7.65]	3.1×10 ⁻⁵	0.19[-0.73: 1.11]	6.8×10 ⁻¹
xanthine	100.0	4.0	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.96[0.69:1.23]	-0.02[-0.09: 0.06]	6.8×10 ⁻¹	0.03[-0.03: 0.09]	3.4×10 ⁻¹	4.07[1.66: 6.49]	9.3×10 ⁻⁴	0.25[-0.69: 1.19]	6.0×10 ⁻¹
X - 11612	87.9	7.0	0.17[0.01:0.32]	0.00[0.00:0.00]	0.83[0.68:0.99]	0.78[0.44:1.11]	-0.01[-0.09: 0.06]	7.0×10 ⁻¹	-0.00[-0.07: 0.06]	9.5×10 ⁻¹	5.24[2.59: 7.89]	1.1×10 ⁻⁴	-0.08[-1.09: 0.94]	8.8×10 ⁻¹
5-oxoproline	99.9	4.4	0.13[0.00:0.28]	0.00[0.00:0.00]	0.87[0.72:1.01]	0.81[0.52:1.11]	0.01[-0.06: 0.09]	7.2×10 ⁻¹	0.01[-0.05: 0.07]	8.4×10 ⁻¹	4.75[2.32: 7.18]	1.2×10 ⁻⁴	0.55[-0.42: 1.53]	2.6×10 ⁻¹
X - 24766	97.8	8.1	0.16[0.02:0.29]	0.00[0.00:0.00]	0.84[0.71:0.98]	0.79[0.49:1.10]	-0.01[-0.09: 0.06]	7.3×10 ⁻¹	0.05[-0.01: 0.11]	1.3×10 ⁻¹	5.45[2.95: 7.95]	1.9×10 ⁻⁵	-0.35[-1.31: 0.61]	4.7×10 ⁻¹
alpha-ketoglutarate	99.4	29.1	0.28[0.14:0.41]	0.00[0.00:0.00]	0.72[0.59:0.86]	0.88[0.60:1.17]	-0.01[-0.09: 0.06]	7.4×10 ⁻¹	0.02[-0.04: 0.08]	5.4×10 ⁻¹	4.19[1.64: 6.73]	1.3×10 ⁻³	0.95[-0.01: 1.91]	5.2×10 ⁻²
propionylglycine	91.1	15.0	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.40[0.05:0.76]	0.01[-0.06: 0.09]	7.5×10 ⁻¹	0.07[0.01: 0.13]	3.2×10 ⁻²	4.79[2.15: 7.43]	3.7×10 ⁻⁴	0.58[-0.42: 1.57]	2.5×10 ⁻¹
serine	100.0	6.4	0.19[0.06:0.33]	0.00[0.00:0.00]	0.81[0.67:0.94]	0.94[0.67:1.21]	-0.01[-0.09: 0.06]	7.7×10 ⁻¹	0.05[-0.01: 0.12]	7.9×10 ⁻²	4.65[2.12: 7.18]	3.2×10 ⁻⁴	0.28[-0.70: 1.27]	5.7×10 ⁻¹
phenyllactate (PLA)	99.7	6.7	0.00[0.00:0.00]	0.15[0.04:0.26]	0.85[0.74:0.96]	0.57[0.24:0.90]	-0.01[-0.09: 0.06]	7.8×10 ⁻¹	0.05[-0.01: 0.12]	8.6×10 ⁻²	4.66[2.18: 7.15]	2.4×10 ⁻⁴	0.14[-0.80: 1.09]	7.7×10 ⁻¹
gamma-glutamylalanine	99.1	5.3	0.24[0.11:0.36]	0.00[0.00:0.00]	0.76[0.64:0.89]	0.86[0.57:1.15]	0.01[-0.07: 0.09]	7.8×10 ⁻¹	-0.04[-0.10: 0.03]	2.6×10 ⁻¹	4.57[2.02: 7.11]	4.4×10 ⁻⁴	0.96[0.02: 1.91]	4.5×10 ⁻²
N-acetylaspargate (NAA)	100.0	19.1	0.00[0.00:0.00]	0.15[0.04:0.26]	0.85[0.74:0.96]	0.79[0.49:1.09]	0.01[-0.06: 0.08]	7.8×10 ⁻¹	-0.01[-0.07: 0.06]	8.6×10 ⁻¹	4.55[2.11: 6.99]	2.6×10 ⁻⁴	0.72[-0.22: 1.67]	1.3×10 ⁻¹
dihomo-linolenate (20:3n3 or n6)	98.7	25.0	0.23[0.09:0.37]	0.00[0.00:0.00]	0.77[0.63:0.91]	0.77[0.46:1.07]	-0.01[-0.08: 0.06]	7.8×10 ⁻¹	0.11[0.04: 0.17]	8.8×10 ⁻⁴	4.47[1.92: 7.01]	5.8×10 ⁻⁴	-0.11[-1.07: 0.85]	8.2×10

2-hydroxybutyrate/2-hydroxyisobutyrate	99.0	8.5	0.00[0.00:0.00]	0.14[0.03:0.26]	0.86[0.74:0.97]	0.82[0.52:1.12]	0.01[-0.07: 0.08]	8.9×10 ⁻¹	0.06[-0.01: 0.12]	7.5×10 ⁻²	4.75[2.23: 7.27]	2.2×10 ⁻⁴	0.22[-0.75: 1.20]	6.5×10 ⁻¹
gamma-glutamylmethionine	98.0	8.3	0.00[0.00:0.00]	0.20[0.08:0.31]	0.80[0.69:0.92]	0.91[0.63:1.19]	0.01[-0.07: 0.08]	9.0×10 ⁻¹	0.03[-0.03: 0.09]	2.9×10 ⁻¹	4.84[2.32: 7.37]	1.7×10 ⁻⁴	0.28[-0.69: 1.25]	5.7×10 ⁻¹
N-acetylasparagine	99.7	10.8	0.00[0.00:0.00]	0.09[0.00:0.20]	0.91[0.80:1.02]	0.50[0.17:0.83]	0.00[-0.07: 0.08]	9.1×10 ⁻¹	0.06[-0.00: 0.12]	5.1×10 ⁻²	5.71[3.17: 8.26]	1.1×10 ⁻⁵	-0.09[-1.07: 0.89]	8.5×10 ⁻¹
choline	100.0	4.6	0.00[0.00:0.00]	0.16[0.05:0.27]	0.84[0.73:0.95]	0.68[0.36:0.99]	-0.00[-0.08: 0.07]	9.2×10 ⁻¹	0.01[-0.05: 0.07]	7.1×10 ⁻¹	4.43[1.96: 6.91]	4.5×10 ⁻⁴	0.65[-0.31: 1.61]	1.8×10 ⁻¹
leucine	100.0	6.0	0.12[0.00:0.26]	0.00[0.00:0.00]	0.88[0.74:1.02]	0.92[0.64:1.20]	0.00[-0.07: 0.08]	9.2×10 ⁻¹	0.04[-0.02: 0.10]	2.5×10 ⁻¹	4.57[2.04: 7.09]	3.9×10 ⁻⁴	0.23[-0.74: 1.21]	6.4×10 ⁻¹
phenylalanine	100.0	5.3	0.17[0.03:0.31]	0.00[0.00:0.00]	0.83[0.69:0.97]	0.98[0.71:1.24]	0.00[-0.07: 0.08]	9.2×10 ⁻¹	0.04[-0.02: 0.10]	2.0×10 ⁻¹	5.37[2.85: 7.89]	3.0×10 ⁻⁵	0.06[-0.92: 1.04]	9.0×10 ⁻¹
uracil	100.0	5.2	0.15[0.00:0.29]	0.00[0.00:0.00]	0.85[0.71:1.00]	1.00[0.75:1.25]	-0.00[-0.08: 0.07]	9.3×10 ⁻¹	0.03[-0.03: 0.09]	3.1×10 ⁻¹	5.02[2.56: 7.49]	6.4×10 ⁻⁵	0.33[-0.61: 1.28]	4.9×10 ⁻¹
N-acetyltyrosine	99.9	6.4	0.00[0.00:0.00]	0.16[0.05:0.28]	0.84[0.72:0.95]	0.79[0.49:1.09]	0.00[-0.07: 0.08]	9.4×10 ⁻¹	0.04[-0.02: 0.10]	2.3×10 ⁻¹	4.45[1.87: 7.02]	7.2×10 ⁻⁴	-0.26[-1.25: 0.73]	6.1×10 ⁻¹
arachidonate (20:4n6)	99.4	18.3	0.13[0.00:0.26]	0.00[0.00:0.00]	0.87[0.74:1.01]	0.77[0.47:1.08]	-0.00[-0.07: 0.07]	9.6×10 ⁻¹	0.13[0.07: 0.19]	1.1×10 ⁻⁵	5.07[2.55: 7.59]	8.2×10 ⁻⁵	-0.33[-1.28: 0.63]	5.0×10 ⁻¹
tyrosine	100.0	6.8	0.00[0.00:0.00]	0.00[0.00:0.00]	1.00[1.00:1.00]	0.92[0.65:1.20]	-0.00[-0.07: 0.07]	9.9×10 ⁻¹	0.06[-0.00: 0.12]	6.1×10 ⁻²	4.64[2.15: 7.12]	2.6×10 ⁻⁴	0.00[-0.98: 0.98]	1.0×10 ⁺⁰

APPENDIX E

Differential multi-fluid networks identify processes involved in end-stage renal disease

Table E.1 Metabolites included in the differential model. I included 882 metabolites – 229 measured in plasma, 419 in urine, and 164 in saliva – in the model. Here I list the results of univariate logistic regression models comparing metabolite abundances between cases and controls (odds ratios and p-values), the node degree (D), betweenness (B), and clustering coefficients (C) from the graphs of controls and cases, respectively. Lastly, the number of differential edges connected to the node is shown.

Metabolite		Association		Controls			Cases			Dif
		OR	p-value	D	B	C	D	B	C	
urine	N2-methylguanosine	0.2 [0.1 : 0.2]	5.9×10^{-32}	8	0.0%	10.7%	1	0.0%	—	0
saliva	undecanedioate	6.9 [4.9 : 9.6]	2.4×10^{-29}	6	0.0%	6.7%	11	0.0%	47.3%	0
urine	sucrose	6.0 [4.4 : 8.2]	2.8×10^{-29}	4	0.0%	16.7%	5	0.0%	0.0%	0
saliva	X - 15689	8.8 [6.0 : 12.9]	8.6×10^{-29}	6	0.0%	20.0%	5	0.0%	30.0%	0
saliva	X - 15675	6.4 [4.6 : 8.9]	1.2×10^{-28}	8	0.0%	25.0%	12	0.0%	16.7%	0
urine	X - 12748	0.2 [0.1 : 0.2]	1.2×10^{-28}	9	0.0%	8.3%	0	0.0%	—	0
urine	X - 12715	8.7 [6.0 : 12.8]	1.3×10^{-28}	4	0.0%	33.3%	0	0.0%	—	0
urine	tetrahydrocortisone	0.2 [0.1 : 0.2]	3.3×10^{-28}	6	0.0%	40.0%	6	0.0%	20.0%	0
plasma	4-hydroxyhippurate	7.3 [5.1 : 10.4]	3.8×10^{-28}	11	0.0%	3.6%	2	0.0%	100.0%	0
urine	X - 17337	0.2 [0.1 : 0.2]	4.9×10^{-28}	12	0.0%	25.8%	5	0.0%	70.0%	0
plasma	pyridoxate	6.2 [4.5 : 8.5]	5.8×10^{-28}	10	0.0%	6.7%	2	0.0%	100.0%	0
saliva	catechol sulfate	9.6 [6.4 : 14.4]	6.0×10^{-28}	9	0.0%	19.4%	4	0.0%	0.0%	0
plasma	X - 19133	0.1 [0.1 : 0.2]	8.7×10^{-28}	11	0.0%	16.4%	7	0.0%	14.3%	0
saliva	azelate (nonanedioate)	7.2 [5.0 : 10.3]	1.2×10^{-27}	7	0.0%	4.8%	9	0.0%	55.6%	0
urine	N-acetylphenylalanine	13.9 [8.7 : 22.3]	1.3×10^{-27}	4	0.0%	50.0%	1	0.0%	—	0
urine	X - 13249	6.6 [4.7 : 9.3]	1.7×10^{-27}	8	0.0%	7.1%	4	0.0%	0.0%	0
saliva	X - 18808	7.0 [5.0 : 10.0]	1.8×10^{-27}	11	0.0%	1.8%	4	0.0%	50.0%	0
urine	X - 17340	0.1 [0.1 : 0.2]	3.2×10^{-27}	10	0.0%	17.8%	6	0.0%	60.0%	0
urine	X - 17341	0.2 [0.2 : 0.3]	3.4×10^{-27}	13	0.0%	19.2%	9	0.0%	44.4%	0
plasma	trigonelline (N'-methylnicotinate)	7.6 [5.3 : 11.0]	3.5×10^{-27}	8	0.0%	10.7%	1	0.0%	—	0
saliva	X - 13007	8.4 [5.7 : 12.4]	3.9×10^{-27}	6	0.0%	26.7%	9	0.0%	30.6%	0
plasma	X - 12846	7.2 [5.0 : 10.3]	4.0×10^{-27}	8	0.0%	3.6%	0	0.0%	—	0
saliva	p-cresol sulfate	5.7 [4.2 : 7.8]	4.5×10^{-27}	11	0.0%	1.8%	5	0.0%	20.0%	1
urine	2-oxo-1-pyrrolidinepropionate	4.3 [3.3 : 5.6]	5.1×10^{-27}	7	0.0%	4.8%	3	0.0%	100.0%	0
plasma	1,5-anhydroglucitol (1,5-AG)	0.1 [0.1 : 0.2]	5.4×10^{-27}	11	0.0%	1.8%	3	0.0%	33.3%	0
plasma	N1-Methyl-2-pyridone-5-carboxamide	14.4 [8.8 : 23.3]	5.9×10^{-27}	5	0.0%	30.0%	5	0.0%	50.0%	0
urine	X - 12831	13.9 [8.6 : 22.5]	6.0×10^{-27}	6	0.0%	26.7%	3	0.0%	0.0%	0
saliva	phenol sulfate	7.8 [5.4 : 11.4]	9.2×10^{-27}	13	0.0%	9.0%	6	0.0%	46.7%	0
urine	X - 17324	0.1 [0.1 : 0.1]	1.1×10^{-26}	11	0.0%	5.5%	0	0.0%	—	0
plasma	X - 11593	8.5 [5.7 : 12.5]	1.1×10^{-26}	7	0.0%	4.8%	1	0.0%	—	0
plasma	hypoxanthine	6.8 [4.8 : 9.7]	1.2×10^{-26}	18	0.0%	6.5%	3	0.0%	0.0%	0
urine	X - 12199	12.1 [7.7 : 19.1]	1.3×10^{-26}	12	0.0%	3.0%	1	0.0%	—	0
saliva	X - 16626	13.5 [8.3 : 21.7]	1.6×10^{-26}	6	0.0%	20.0%	5	0.0%	50.0%	0
saliva	aspartate	0.2 [0.2 : 0.3]	1.6×10^{-26}	9	0.0%	19.4%	9	0.0%	33.3%	0
urine	X - 12834	5.7 [4.1 : 7.9]	1.7×10^{-26}	11	0.0%	10.9%	1	0.0%	—	0
urine	X - 10445	0.1 [0.1 : 0.2]	1.8×10^{-26}	10	0.0%	0.0%	4	0.0%	0.0%	0
urine	X - 13844	16.8 [10.0 : 28.2]	1.9×10^{-26}	7	0.0%	19.0%	2	0.0%	100.0%	0
urine	4-acetamidobutanoate	11.6 [7.4 : 18.2]	2.4×10^{-26}	9	0.0%	2.8%	1	0.0%	—	0
urine	X - 20180	6.5 [4.6 : 9.1]	3.2×10^{-26}	11	0.0%	3.6%	3	0.0%	0.0%	0
urine	5-oxoproline	6.0 [4.3 : 8.4]	3.7×10^{-26}	16	0.0%	2.5%	2	0.0%	100.0%	0
plasma	catechol sulfate	8.9 [5.9 : 13.4]	3.8×10^{-26}	6	0.0%	13.3%	2	0.0%	0.0%	0

continued on next page ...

urine	homovanillate sulfate	7.5[5.1: 10.8]	4.1×10^{-26}	14	0.0%	2.2%	1	0.0%	—	0
urine	X - 12379	7.1[4.9: 10.2]	4.4×10^{-26}	11	0.0%	14.5%	1	0.0%	—	0
urine	N6-acetyllysine	0.1[0.0: 0.1]	4.7×10^{-26}	12	0.0%	4.5%	1	0.0%	—	0
plasma	X - 11793	0.1[0.1: 0.2]	5.2×10^{-26}	6	0.0%	20.0%	4	0.0%	16.7%	0
plasma	X - 12095	9.5[6.3: 14.5]	5.5×10^{-26}	5	0.0%	30.0%	6	0.0%	46.7%	0
urine	X - 13684	0.1[0.1: 0.2]	5.8×10^{-26}	13	0.0%	3.8%	2	0.0%	100.0%	0
urine	X - 12814	14.3[8.7: 23.5]	5.9×10^{-26}	7	0.0%	9.5%	1	0.0%	—	0
plasma	X - 16397	14.8[8.9: 24.4]	6.2×10^{-26}	15	0.0%	1.0%	0	0.0%	—	0
urine	X - 12231	0.2[0.2: 0.3]	7.1×10^{-26}	6	0.0%	86.7%	6	0.0%	66.7%	0
plasma	X - 11442	0.1[0.1: 0.2]	7.4×10^{-26}	8	0.0%	67.9%	8	0.0%	60.7%	0
plasma	kynurenine	8.5[5.7: 12.7]	7.6×10^{-26}	11	0.0%	10.9%	0	0.0%	—	0
saliva	X - 18554	6.2[4.4: 8.6]	7.9×10^{-26}	7	0.0%	4.8%	10	0.0%	35.6%	0
urine	2-isopropylmalate	3.8[2.9: 4.8]	8.3×10^{-26}	6	0.0%	6.7%	4	0.0%	33.3%	0
urine	X - 11485	0.1[0.1: 0.2]	8.4×10^{-26}	10	0.0%	31.1%	7	0.0%	61.9%	0
plasma	gamma-glutamylvaline	9.2[6.1: 13.9]	8.6×10^{-26}	10	0.0%	24.4%	5	0.0%	50.0%	0
plasma	succinylcarnitine	10.1[6.6: 15.6]	9.1×10^{-26}	14	0.0%	1.1%	0	0.0%	—	0
urine	X - 12844	0.2[0.1: 0.3]	9.5×10^{-26}	7	0.0%	28.6%	10	0.0%	44.4%	1
plasma	p-cresol sulfate	12.8[7.9: 20.6]	1.2×10^{-25}	4	0.0%	16.7%	6	0.0%	46.7%	1
urine	X - 14445	7.9[5.3: 11.6]	1.5×10^{-25}	13	0.0%	9.0%	2	0.0%	0.0%	0
plasma	indolelactate	7.0[4.9: 10.1]	1.5×10^{-25}	11	0.0%	10.9%	3	0.0%	0.0%	0
plasma	hippurate	6.5[4.5: 9.2]	1.6×10^{-25}	9	0.0%	27.8%	1	0.0%	—	0
plasma	3-methylglutarylarnitine (C6)	18.5[10.7: 32.0]	1.6×10^{-25}	11	0.0%	16.4%	1	0.0%	—	0
urine	4-androsten-3beta,17beta-diol disulfate (2)	0.1[0.1: 0.2]	1.7×10^{-25}	10	0.0%	35.6%	6	0.0%	33.3%	0
saliva	N1-Methyl-2-pyridone-5-carboxamide	17.2[10.1: 29.3]	1.7×10^{-25}	10	0.0%	13.3%	8	0.0%	17.9%	0
urine	X - 12848	0.2[0.1: 0.2]	1.8×10^{-25}	7	0.0%	52.4%	5	0.0%	90.0%	0
plasma	X - 16934	0.2[0.1: 0.2]	2.1×10^{-25}	6	0.0%	80.0%	7	0.0%	76.2%	0
plasma	glutarylarnitine (C5)	10.8[6.9: 17.0]	2.1×10^{-25}	2	0.0%	0.0%	1	0.0%	—	0
urine	X - 12733	4.9[3.6: 6.6]	2.3×10^{-25}	8	0.0%	17.9%	2	0.0%	0.0%	0
urine	5beta-pregnan-3alpha,21-diol-11,20-dione 21-glucosiduronate	0.1[0.0: 0.1]	2.3×10^{-25}	9	0.0%	33.3%	5	0.0%	10.0%	0
plasma	X - 19141	10.4[6.7: 16.3]	2.6×10^{-25}	12	0.0%	3.0%	0	0.0%	—	0
plasma	X - 11378	0.1[0.1: 0.1]	2.8×10^{-25}	7	0.0%	14.3%	11	0.0%	12.7%	0
plasma	3-hydroxyhippurate	7.1[4.9: 10.2]	3.7×10^{-25}	14	0.0%	5.5%	1	0.0%	—	0
urine	dihydrobiopterin	0.2[0.2: 0.3]	3.8×10^{-25}	15	0.0%	3.8%	0	0.0%	—	0
plasma	X - 12206	7.9[5.3: 11.7]	3.9×10^{-25}	16	0.0%	4.2%	5	0.0%	10.0%	0
urine	uracil	0.2[0.1: 0.3]	4.0×10^{-25}	8	0.0%	3.6%	1	0.0%	—	0
plasma	X - 17357	7.8[5.3: 11.6]	4.5×10^{-25}	8	0.0%	14.3%	2	0.0%	0.0%	0
saliva	X - 16271	4.9[3.7: 6.7]	4.6×10^{-25}	13	0.0%	2.6%	9	0.0%	38.9%	0
saliva	X - 13671	7.0[4.9: 10.2]	4.6×10^{-25}	6	0.0%	46.7%	8	0.0%	35.7%	2
plasma	gamma-glutamylphenylalanine	9.7[6.3: 15.0]	4.9×10^{-25}	7	0.0%	52.4%	5	0.0%	30.0%	0
urine	X - 12714	5.0[3.7: 6.8]	5.0×10^{-25}	9	0.0%	5.6%	5	0.0%	10.0%	0
urine	4-hydroxyhippurate	4.2[3.2: 5.6]	5.0×10^{-25}	5	0.0%	20.0%	8	0.0%	17.9%	1
plasma	X - 12101	12.4[7.7: 20.0]	5.5×10^{-25}	14	0.0%	0.0%	2	0.0%	100.0%	0
plasma	phosphate	16.9[9.9: 28.9]	5.8×10^{-25}	12	0.0%	7.6%	4	0.0%	16.7%	0
plasma	10-undecenoate (11:1n1)	0.2[0.1: 0.3]	5.8×10^{-25}	12	0.0%	12.1%	0	0.0%	—	0
urine	X - 17327	6.1[4.3: 8.6]	6.5×10^{-25}	9	0.0%	30.6%	4	0.0%	33.3%	0
urine	hypoxanthine	0.1[0.0: 0.1]	6.6×10^{-25}	18	0.0%	7.8%	9	0.0%	22.2%	1
saliva	acisoga	8.5[5.7: 12.8]	7.3×10^{-25}	13	0.0%	14.1%	10	0.0%	8.9%	0
plasma	X - 17359	9.1[6.0: 13.9]	7.5×10^{-25}	11	0.0%	3.6%	0	0.0%	—	0
urine	3-sialyllactose	12.3[7.6: 19.9]	7.7×10^{-25}	13	0.0%	3.8%	6	0.0%	6.7%	0
plasma	tryptophan	0.1[0.1: 0.2]	8.6×10^{-25}	14	0.0%	6.6%	1	0.0%	—	0
urine	N-acetyltyrosine	4.9[3.6: 6.7]	1.1×10^{-24}	11	0.0%	9.1%	7	0.0%	4.8%	0
urine	C-glycosyltryptophan	16.7[9.7: 28.5]	1.2×10^{-24}	16	0.0%	3.3%	3	0.0%	0.0%	0
saliva	X - 18111	24.9[13.5: 46.0]	1.3×10^{-24}	6	0.0%	53.3%	5	0.0%	30.0%	0
urine	X - 12906	7.2[4.9: 10.5]	1.4×10^{-24}	13	0.0%	0.0%	1	0.0%	—	0
plasma	biliverdin	0.2[0.1: 0.2]	1.4×10^{-24}	8	0.0%	60.7%	6	0.0%	66.7%	0
plasma	3-indoxyl sulfate	9.7[6.3: 14.9]	1.5×10^{-24}	8	0.0%	21.4%	1	0.0%	—	0
urine	X - 20574	6.5[4.6: 9.3]	1.5×10^{-24}	7	0.0%	28.6%	2	0.0%	0.0%	0
plasma	glycerophosphorylcholine (GPC)	0.1[0.0: 0.1]	1.5×10^{-24}	17	0.0%	7.4%	9	0.0%	5.6%	0
urine	adenosine 3',5'-cyclic monophosphate (cAMP)	0.1[0.1: 0.2]	1.5×10^{-24}	13	0.0%	5.1%	1	0.0%	—	0
urine	X - 20571	7.2[4.9: 10.6]	1.7×10^{-24}	11	0.0%	0.0%	6	0.0%	6.7%	0
plasma	2-methylbutyrylcarnitine (C5)	7.4[5.0: 10.9]	1.7×10^{-24}	8	0.0%	7.1%	6	0.0%	66.7%	0
plasma	choline	6.1[4.3: 8.6]	1.9×10^{-24}	14	0.0%	2.2%	1	0.0%	—	0
urine	X - 11452	0.2[0.2: 0.3]	1.9×10^{-24}	7	0.0%	76.2%	6	0.0%	80.0%	0
plasma	N-formylmethionine	7.2[4.9: 10.5]	2.1×10^{-24}	18	0.0%	2.6%	1	0.0%	—	0
urine	X - 11945	12.3[7.6: 19.9]	2.3×10^{-24}	8	0.0%	21.4%	1	0.0%	—	0
plasma	X - 11315	0.1[0.1: 0.2]	2.3×10^{-24}	8	0.0%	0.0%	0	0.0%	—	0
plasma	N-acetylphenylalanine	9.9[6.4: 15.4]	2.5×10^{-24}	16	0.0%	0.8%	1	0.0%	—	0
plasma	urea	26.9[14.3: 50.7]	2.8×10^{-24}	10	0.0%	8.9%	0	0.0%	—	1
plasma	isobutyrylcarnitine	6.9[4.7: 10.0]	3.1×10^{-24}	13	0.0%	6.4%	6	0.0%	53.3%	0
urine	N1-methyladenosine	0.2[0.1: 0.2]	3.9×10^{-24}	9	0.0%	13.9%	0	0.0%	—	0
plasma	4-methyl-2-oxopentanoate	0.1[0.1: 0.2]	4.0×10^{-24}	6	0.0%	40.0%	2	0.0%	100.0%	0
urine	X - 12794	21.1[11.7: 38.0]	4.1×10^{-24}	19	0.0%	1.2%	1	0.0%	—	0
urine	beta-hydroxyisovalerate	0.1[0.0: 0.1]	4.8×10^{-24}	9	0.0%	5.6%	2	0.0%	0.0%	0
plasma	3-methyl-2-oxobutyrate	0.2[0.1: 0.3]	5.1×10^{-24}	10	0.0%	11.1%	3	0.0%	33.3%	0
saliva	threonine	0.2[0.2: 0.3]	5.2×10^{-24}	17	0.0%	9.6%	12	0.0%	21.2%	0
plasma	X - 11441	0.2[0.1: 0.3]	6.7×10^{-24}	6	0.0%	60.0%	8	0.0%	60.7%	0
urine	imidazole lactate	0.1[0.1: 0.2]	7.6×10^{-24}	15	0.0%	6.7%	5	0.0%	10.0%	0
plasma	cortisone	0.2[0.1: 0.3]	7.7×10^{-24}	9	0.0%	5.6%	15	0.0%	0.0%	0
plasma	gamma-glutamylisoleucine	27.9[14.6: 53.3]	7.9×10^{-24}	9	0.0%	19.4%	2	0.0%	100.0%	0

continued on next page ...

urine	orotidine	3.0[2.4: 3.7]	1.1×10^{-23}	15	0.0%	13.3%	11	0.0%	30.9%	0
plasma	tiglyl carnitine	7.4[5.0: 11.0]	1.2×10^{-23}	9	0.0%	2.8%	6	0.0%	33.3%	0
plasma	N-acetyl-beta-alanine	3.8[2.9: 4.9]	1.7×10^{-23}	9	0.0%	8.3%	5	0.0%	20.0%	0
urine	N1-methylguanosine	0.1[0.1: 0.2]	2.1×10^{-23}	18	0.0%	5.2%	0	0.0%	—	0
plasma	N-acetylalanine	24.8[13.2: 46.7]	2.2×10^{-23}	9	0.0%	2.8%	0	0.0%	—	0
urine	X - 12739	4.2[3.1: 5.5]	2.7×10^{-23}	8	0.0%	35.7%	3	0.0%	66.7%	0
urine	X - 12329	3.6[2.8: 4.6]	3.0×10^{-23}	7	0.0%	38.1%	5	0.0%	30.0%	1
saliva	X - 19555	31.4[15.9: 62.1]	3.3×10^{-23}	8	0.0%	10.7%	2	0.0%	100.0%	0
urine	X - 17717	4.0[3.0: 5.2]	3.3×10^{-23}	8	0.0%	28.6%	5	0.0%	50.0%	0
saliva	X - 14904	3.9[3.0: 5.1]	3.4×10^{-23}	7	0.0%	9.5%	4	0.0%	33.3%	0
plasma	phenol sulfate	3.8[2.9: 4.9]	3.5×10^{-23}	5	0.0%	40.0%	5	0.0%	70.0%	0
urine	X - 20586	6.5[4.5: 9.4]	4.1×10^{-23}	9	0.0%	11.1%	0	0.0%	—	0
urine	X - 12690	6.9[4.7: 10.2]	4.3×10^{-23}	12	0.0%	10.6%	2	0.0%	0.0%	0
plasma	O-sulfo-L-tyrosine	21.8[11.8: 40.2]	5.3×10^{-23}	4	0.0%	33.3%	0	0.0%	—	0
urine	X - 13726	4.6[3.4: 6.2]	5.3×10^{-23}	3	0.0%	66.7%	4	0.0%	100.0%	0
plasma	X - 11880	0.2[0.1: 0.2]	5.5×10^{-23}	4	0.0%	50.0%	5	0.0%	60.0%	0
urine	X - 19779	4.6[3.4: 6.3]	5.6×10^{-23}	12	0.0%	4.5%	6	0.0%	26.7%	0
plasma	X - 11444	5.2[3.7: 7.2]	5.7×10^{-23}	10	0.0%	24.4%	3	0.0%	33.3%	0
saliva	X - 18113	25.5[13.4: 48.6]	5.9×10^{-23}	6	0.0%	40.0%	5	0.0%	40.0%	0
plasma	2-hydroxystearate	0.2[0.2: 0.3]	6.1×10^{-23}	2	0.0%	0.0%	1	0.0%	—	0
plasma	X - 11372	0.1[0.1: 0.2]	7.4×10^{-23}	4	0.0%	50.0%	6	0.0%	53.3%	0
plasma	dihomo-linolenate (20:3n3 or n6)	0.2[0.2: 0.3]	7.5×10^{-23}	10	0.0%	22.2%	4	0.0%	50.0%	0
urine	X - 17323	0.2[0.2: 0.3]	7.6×10^{-23}	15	0.0%	4.8%	0	0.0%	—	0
urine	N-(2-furoyl)glycine	3.4[2.7: 4.4]	7.8×10^{-23}	9	0.0%	27.8%	2	0.0%	0.0%	0
urine	X - 12704	6.4[4.4: 9.2]	9.7×10^{-23}	6	0.0%	40.0%	4	0.0%	83.3%	0
urine	1-methylimidazoleacetate	4.0[3.0: 5.2]	1.1×10^{-22}	14	0.0%	4.4%	3	0.0%	0.0%	0
urine	ferulic acid 4-sulfate	4.4[3.3: 6.0]	1.1×10^{-22}	8	0.0%	14.3%	8	0.0%	3.6%	0
saliva	glutamine	0.2[0.2: 0.3]	1.2×10^{-22}	12	0.0%	15.2%	10	0.0%	28.9%	0
saliva	8-hydroxyoctanoate	3.8[2.9: 4.9]	1.3×10^{-22}	8	0.0%	14.3%	10	0.0%	46.7%	0
urine	neopterin	4.9[3.5: 6.7]	1.5×10^{-22}	13	0.0%	3.8%	4	0.0%	0.0%	0
plasma	arachidonate (20:4n6)	0.2[0.2: 0.3]	1.5×10^{-22}	12	0.0%	31.8%	4	0.0%	33.3%	0
urine	X - 13836	5.2[3.7: 7.3]	1.5×10^{-22}	6	0.0%	6.7%	1	0.0%	—	0
saliva	X - 18672	0.2[0.2: 0.3]	1.9×10^{-22}	13	0.0%	7.7%	16	0.0%	5.8%	0
plasma	X - 11564	27.7[14.2: 54.1]	1.9×10^{-22}	12	0.0%	6.1%	0	0.0%	—	0
urine	X - 15363	4.7[3.4: 6.4]	2.4×10^{-22}	10	0.0%	6.7%	2	0.0%	100.0%	0
plasma	1-methyl-2-piperidinecarboxylic acid	33.7[16.6: 68.4]	2.5×10^{-22}	5	0.0%	40.0%	7	0.0%	4.8%	1
urine	2-methylbutyrylcarnitine (C5)	0.3[0.2: 0.4]	3.8×10^{-22}	10	0.0%	22.2%	3	0.0%	33.3%	0
urine	X - 17683	3.7[2.8: 4.8]	5.5×10^{-22}	15	0.0%	1.9%	1	0.0%	—	0
urine	gamma-glutamylvaline	0.2[0.2: 0.3]	5.5×10^{-22}	7	0.0%	47.6%	6	0.0%	66.7%	0
urine	X - 20598	4.9[3.6: 6.8]	5.7×10^{-22}	5	0.0%	30.0%	3	0.0%	33.3%	0
urine	3-methylcytidine	0.1[0.1: 0.1]	6.6×10^{-22}	11	0.0%	7.3%	0	0.0%	—	0
urine	citrate	0.3[0.2: 0.3]	8.4×10^{-22}	18	0.0%	4.6%	3	0.0%	33.3%	0
urine	tiglyl carnitine	0.2[0.2: 0.3]	8.7×10^{-22}	6	0.0%	0.0%	1	0.0%	—	0
plasma	pro-hydroxy-pro	31.3[15.5: 63.5]	1.1×10^{-21}	10	0.0%	13.3%	3	0.0%	66.7%	0
urine	N6-carbamoylthreonyladenosine	4.7[3.4: 6.5]	1.3×10^{-21}	6	0.0%	40.0%	3	0.0%	0.0%	0
saliva	X - 20849	38.2[18.1: 81.0]	1.6×10^{-21}	5	0.0%	50.0%	4	0.0%	33.3%	0
urine	X - 12113	4.0[3.0: 5.3]	2.2×10^{-21}	16	0.0%	10.8%	4	0.0%	0.0%	0
plasma	N-acetylthreonine	29.9[14.8: 60.3]	2.5×10^{-21}	9	0.0%	2.8%	0	0.0%	—	0
urine	X - 18367	3.2[2.5: 4.0]	3.6×10^{-21}	7	0.0%	47.6%	6	0.0%	60.0%	0
urine	X - 17335	0.0[0.0: 0.1]	3.9×10^{-21}	10	0.0%	13.3%	0	0.0%	—	0
plasma	eicosapentaenoate (EPA; 20:5n3)	0.2[0.2: 0.3]	3.9×10^{-21}	13	0.0%	17.9%	5	0.0%	60.0%	0
urine	androsterone sulfate	0.2[0.2: 0.3]	4.0×10^{-21}	8	0.0%	35.7%	8	0.0%	28.6%	0
plasma	carnitine	0.2[0.1: 0.3]	4.1×10^{-21}	14	0.0%	4.4%	1	0.0%	—	0
saliva	arginine	0.3[0.2: 0.4]	4.8×10^{-21}	15	0.0%	9.5%	5	0.0%	20.0%	0
urine	X - 20590	4.8[3.5: 6.7]	5.6×10^{-21}	16	0.0%	5.8%	4	0.0%	0.0%	0
plasma	5-oxoproline	4.2[3.1: 5.6]	5.7×10^{-21}	13	0.0%	5.1%	1	0.0%	—	0
urine	X - 13709	4.4[3.2: 6.0]	6.5×10^{-21}	7	0.0%	19.0%	3	0.0%	33.3%	0
plasma	2-stearoylglycerophosphocholine	0.3[0.2: 0.3]	7.5×10^{-21}	4	0.0%	16.7%	7	0.0%	19.0%	0
urine	3-hydroxy-2-ethylpropionate	0.3[0.2: 0.4]	7.6×10^{-21}	10	0.0%	6.7%	2	0.0%	0.0%	0
plasma	phenylacetylglutamine	40.1[18.5: 87.0]	9.8×10^{-21}	7	0.0%	19.0%	0	0.0%	—	0
urine	X - 15503	4.6[3.3: 6.4]	1.1×10^{-20}	6	0.0%	13.3%	0	0.0%	—	0
urine	7-methylguanine	0.2[0.2: 0.3]	1.1×10^{-20}	14	0.0%	11.0%	0	0.0%	—	0
urine	X - 15812	3.7[2.8: 4.8]	1.1×10^{-20}	12	0.0%	0.0%	2	0.0%	0.0%	0
urine	hydroquinone sulfate	4.1[3.1: 5.6]	1.2×10^{-20}	11	0.0%	10.9%	7	0.0%	42.9%	0
saliva	X - 19703	41.5[18.9: 90.9]	1.3×10^{-20}	9	0.0%	13.9%	3	0.0%	33.3%	0
urine	N-acetylarginine	0.2[0.2: 0.3]	1.4×10^{-20}	14	0.0%	17.6%	4	0.0%	33.3%	0
plasma	pantothenate	4.0[3.0: 5.4]	1.6×10^{-20}	12	0.0%	4.5%	7	0.0%	23.8%	0
plasma	X - 11308	0.2[0.2: 0.3]	1.8×10^{-20}	8	0.0%	10.7%	5	0.0%	60.0%	0
plasma	deoxycarnitine	4.0[3.0: 5.4]	2.3×10^{-20}	10	0.0%	2.2%	4	0.0%	0.0%	0
urine	pyroglutamylvaline	4.1[3.1: 5.6]	2.6×10^{-20}	11	0.0%	7.3%	5	0.0%	10.0%	0
urine	glycocholate sulfate	0.2[0.2: 0.3]	2.7×10^{-20}	11	0.0%	7.3%	3	0.0%	33.3%	0
plasma	2-hydroxypalmitate	0.3[0.2: 0.3]	3.2×10^{-20}	7	0.0%	4.8%	2	0.0%	0.0%	0
plasma	X - 12096	32.8[15.6: 68.9]	3.7×10^{-20}	14	0.0%	8.8%	0	0.0%	—	0
plasma	docosahexaenoate (DHA; 22:6n3)	0.2[0.2: 0.3]	3.8×10^{-20}	10	0.0%	28.9%	6	0.0%	40.0%	0
plasma	1-dihomo-linoleoylglycerophosphocholine (20:2n6)	0.3[0.2: 0.3]	4.6×10^{-20}	5	0.0%	10.0%	5	0.0%	80.0%	0
plasma	X - 11530	0.3[0.2: 0.4]	5.7×10^{-20}	14	0.0%	22.0%	7	0.0%	47.6%	0
plasma	isovalerate	0.3[0.2: 0.4]	5.8×10^{-20}	10	0.0%	8.9%	3	0.0%	33.3%	0
urine	X - 17320	2.7[2.2: 3.4]	5.9×10^{-20}	7	0.0%	33.3%	1	0.0%	—	1
urine	tryptophan	0.2[0.2: 0.3]	6.3×10^{-20}	7	0.0%	23.8%	6	0.0%	26.7%	0

continued on next page ...

plasma	3-methyl-2-oxovalerate	0.3[0.2: 0.3]	9.8×10^{-20}	6	0.0%	33.3%	3	0.0%	33.3%	0
urine	X - 12742	4.3[3.1: 5.9]	1.0×10^{-19}	12	0.0%	24.2%	3	0.0%	66.7%	0
urine	X - 17346	3.1[2.4: 4.0]	1.1×10^{-19}	10	0.0%	0.0%	0	0.0%	—	0
plasma	dehydroisoandrosterone sulfate (DHEA-S)	0.2[0.2: 0.3]	1.1×10^{-19}	12	0.0%	34.8%	9	0.0%	41.7%	0
urine	N-acetylneuraminate	4.0[3.0: 5.5]	1.2×10^{-19}	6	0.0%	6.7%	2	0.0%	0.0%	0
urine	X - 17339	0.3[0.2: 0.4]	1.3×10^{-19}	14	0.0%	12.1%	7	0.0%	81.0%	1
plasma	N4-acetylcytidine	47.2[20.4:109.0]	2.0×10^{-19}	9	0.0%	0.0%	1	0.0%	—	0
plasma	4-acetamidobutanoate	44.9[19.6:102.8]	2.2×10^{-19}	7	0.0%	4.8%	0	0.0%	—	0
saliva	alpha-hydroxyisovalerate	3.3[2.5: 4.2]	2.3×10^{-19}	15	0.0%	7.6%	3	0.0%	0.0%	0
urine	X - 19910	3.4[2.6: 4.5]	2.4×10^{-19}	11	0.0%	3.6%	1	0.0%	—	0
plasma	X - 15461	3.6[2.7: 4.7]	2.6×10^{-19}	18	0.0%	0.7%	5	0.0%	20.0%	0
urine	glycylproline	0.2[0.2: 0.3]	2.8×10^{-19}	6	0.0%	20.0%	9	0.0%	0.0%	0
urine	21-hydroxypregnenolone disulfate	0.2[0.1: 0.3]	3.3×10^{-19}	9	0.0%	25.0%	4	0.0%	33.3%	1
plasma	pseudouridine	43.5[19.1: 99.5]	3.4×10^{-19}	6	0.0%	20.0%	0	0.0%	—	0
plasma	3-methylhistidine	3.6[2.7: 4.8]	3.5×10^{-19}	6	0.0%	20.0%	7	0.0%	14.3%	0
urine	acisoga	0.3[0.2: 0.4]	3.8×10^{-19}	9	0.0%	5.6%	4	0.0%	50.0%	0
urine	X - 12411	3.6[2.7: 4.8]	3.9×10^{-19}	11	0.0%	3.6%	4	0.0%	0.0%	0
urine	dimethylarginine (SDMA + ADMA)	0.1[0.1: 0.2]	4.1×10^{-19}	14	0.0%	8.8%	9	0.0%	13.9%	0
urine	X - 16947	2.8[2.2: 3.5]	6.7×10^{-19}	17	0.0%	0.7%	7	0.0%	9.5%	1
plasma	betaine	0.3[0.2: 0.4]	7.2×10^{-19}	14	0.0%	3.3%	0	0.0%	—	0
urine	X - 12680	3.3[2.6: 4.3]	7.9×10^{-19}	8	0.0%	7.1%	9	0.0%	13.9%	1
urine	X - 13837	3.9[2.9: 5.2]	7.9×10^{-19}	8	0.0%	42.9%	5	0.0%	20.0%	0
urine	N-acetylisoleucine	0.3[0.2: 0.4]	8.2×10^{-19}	12	0.0%	6.1%	8	0.0%	7.1%	0
plasma	X - 11429	46.8[20.0:109.8]	9.0×10^{-19}	5	0.0%	30.0%	0	0.0%	—	0
urine	X - 13698	0.3[0.2: 0.3]	9.2×10^{-19}	10	0.0%	35.6%	7	0.0%	28.6%	0
urine	X - 11261	0.4[0.3: 0.4]	1.1×10^{-18}	10	0.0%	28.9%	3	0.0%	66.7%	0
plasma	2-palmitoylglycerophosphocholine	0.3[0.2: 0.4]	1.1×10^{-18}	7	0.0%	47.6%	9	0.0%	50.0%	0
urine	X - 12565	3.0[2.4: 3.9]	1.2×10^{-18}	13	0.0%	7.7%	2	0.0%	0.0%	0
plasma	1-myristoylglycerophosphocholine (14:0)	0.3[0.2: 0.4]	1.6×10^{-18}	8	0.0%	50.0%	7	0.0%	66.7%	0
urine	X - 17693	3.4[2.6: 4.4]	1.6×10^{-18}	12	0.0%	21.2%	1	0.0%	—	0
plasma	citulline	3.9[2.9: 5.3]	1.7×10^{-18}	13	0.0%	5.1%	0	0.0%	—	0
plasma	1-eicosatrienoylglycerophosphocholine (20:3)	0.3[0.2: 0.4]	1.8×10^{-18}	9	0.0%	25.0%	9	0.0%	47.2%	0
saliva	X - 13537	3.6[2.7: 4.8]	1.8×10^{-18}	5	0.0%	50.0%	6	0.0%	53.3%	0
urine	3-methyladipate	4.2[3.1: 5.8]	1.8×10^{-18}	9	0.0%	0.0%	4	0.0%	16.7%	0
urine	N4-acetylcytidine	0.3[0.3: 0.4]	2.0×10^{-18}	14	0.0%	6.6%	1	0.0%	—	0
saliva	adenine	0.3[0.3: 0.4]	2.0×10^{-18}	19	0.0%	2.9%	9	0.0%	8.3%	1
plasma	tyrosine	0.3[0.2: 0.4]	2.1×10^{-18}	10	0.0%	13.3%	4	0.0%	16.7%	0
urine	X - 11440	0.3[0.2: 0.4]	2.1×10^{-18}	5	0.0%	40.0%	4	0.0%	33.3%	0
urine	tyrosine	0.3[0.2: 0.4]	2.4×10^{-18}	5	0.0%	30.0%	8	0.0%	46.4%	1
plasma	X - 12844	3.6[2.7: 4.8]	2.9×10^{-18}	6	0.0%	26.7%	4	0.0%	33.3%	0
plasma	X - 16580	40.3[17.6: 92.6]	3.0×10^{-18}	15	0.0%	1.9%	0	0.0%	—	0
urine	andro steroid monosulfate 2	0.3[0.2: 0.4]	4.0×10^{-18}	8	0.0%	53.6%	2	0.0%	0.0%	0
plasma	creatine	0.3[0.2: 0.4]	4.1×10^{-18}	14	0.0%	9.9%	11	0.0%	1.8%	1
urine	X - 12760	3.9[2.9: 5.3]	4.3×10^{-18}	13	0.0%	9.0%	2	0.0%	100.0%	0
urine	kynurenate	4.0[2.9: 5.5]	4.9×10^{-18}	14	0.0%	7.7%	11	0.0%	5.5%	0
urine	isobutyrylcarnitine	0.3[0.3: 0.4]	5.9×10^{-18}	8	0.0%	17.9%	1	0.0%	—	0
urine	5-hydroxyindoleacetate	3.5[2.6: 4.7]	6.4×10^{-18}	14	0.0%	1.1%	0	0.0%	—	0
urine	X - 20578	3.7[2.7: 5.0]	6.5×10^{-18}	9	0.0%	33.3%	4	0.0%	50.0%	0
urine	etiocholanolone glucuronide	0.3[0.2: 0.4]	7.0×10^{-18}	18	0.0%	9.8%	8	0.0%	14.3%	0
saliva	13-HODE + 9-HODE	3.3[2.5: 4.3]	7.4×10^{-18}	6	0.0%	53.3%	9	0.0%	33.3%	0
urine	gamma-glutamylisoleucine	0.3[0.2: 0.4]	8.6×10^{-18}	10	0.0%	17.8%	4	0.0%	100.0%	0
plasma	1-palmitoleoylglycerophosphocholine (16:1)	0.3[0.2: 0.4]	8.8×10^{-18}	9	0.0%	36.1%	9	0.0%	63.9%	0
saliva	succinate	3.1[2.4: 4.0]	1.1×10^{-17}	18	0.0%	5.2%	10	0.0%	13.3%	0
saliva	glutamate	0.3[0.2: 0.4]	1.1×10^{-17}	15	0.0%	8.6%	11	0.0%	25.5%	0
urine	pregnen-diol disulfate	0.2[0.2: 0.3]	1.1×10^{-17}	8	0.0%	46.4%	5	0.0%	50.0%	1
urine	4-hydroxyphenylpyruvate	3.3[2.5: 4.4]	1.2×10^{-17}	15	0.0%	5.7%	5	0.0%	10.0%	0
plasma	acisoga	3.4[2.6: 4.5]	1.2×10^{-17}	4	0.0%	33.3%	9	0.0%	13.9%	0
urine	X - 01911	0.3[0.3: 0.4]	1.3×10^{-17}	12	0.0%	33.3%	7	0.0%	42.9%	0
plasma	X - 18929	0.3[0.2: 0.4]	1.8×10^{-17}	13	0.0%	25.6%	6	0.0%	93.3%	0
plasma	1-margaroylglycerophosphocholine (17:0)	0.3[0.2: 0.4]	2.0×10^{-17}	5	0.0%	20.0%	7	0.0%	33.3%	0
saliva	creatinine	2.6[2.1: 3.2]	2.8×10^{-17}	16	0.0%	9.2%	0	0.0%	—	0
urine	adenosine	0.3[0.2: 0.4]	3.0×10^{-17}	15	0.0%	5.7%	11	0.0%	7.3%	0
plasma	4-methylcatechol sulfate	3.3[2.5: 4.3]	3.1×10^{-17}	3	0.0%	33.3%	6	0.0%	46.7%	1
urine	N-acetyl-3-methylhistidine	3.1[2.4: 4.1]	3.3×10^{-17}	6	0.0%	33.3%	6	0.0%	33.3%	0
plasma	N2,N2-dimethylguanosine	55.4[21.8:141.1]	3.6×10^{-17}	15	0.0%	2.9%	0	0.0%	—	0
plasma	X - 16946	0.3[0.2: 0.4]	4.6×10^{-17}	10	0.0%	35.6%	6	0.0%	93.3%	0
urine	X - 17705	0.3[0.2: 0.4]	4.9×10^{-17}	11	0.0%	1.8%	7	0.0%	0.0%	0
urine	X - 12410	3.6[2.7: 4.9]	7.3×10^{-17}	15	0.0%	1.0%	10	0.0%	8.9%	0
urine	lactate	2.6[2.1: 3.3]	8.0×10^{-17}	11	0.0%	1.8%	0	0.0%	—	0
urine	andro steroid monosulfate (1)	0.3[0.2: 0.4]	1.3×10^{-16}	8	0.0%	21.4%	4	0.0%	16.7%	0
urine	X - 12096	3.7[2.7: 5.1]	1.3×10^{-16}	6	0.0%	33.3%	3	0.0%	33.3%	0
plasma	1-palmitoylglycerophosphocholine (16:0)	0.4[0.3: 0.5]	1.5×10^{-16}	5	0.0%	50.0%	13	0.0%	42.3%	0
plasma	N-delta-acetylornithine	3.2[2.4: 4.2]	1.5×10^{-16}	6	0.0%	6.7%	6	0.0%	0.0%	1
plasma	citrate	0.3[0.2: 0.4]	1.7×10^{-16}	10	0.0%	0.0%	2	0.0%	0.0%	0
urine	azelate (nonanedioate)	2.6[2.1: 3.3]	1.7×10^{-16}	13	0.0%	6.4%	4	0.0%	0.0%	0
urine	X - 12234	2.5[2.0: 3.2]	1.7×10^{-16}	6	0.0%	80.0%	7	0.0%	66.7%	0
urine	N-acetylparagine	0.2[0.2: 0.3]	1.7×10^{-16}	9	0.0%	38.9%	4	0.0%	50.0%	0
plasma	stearate (18:0)	0.4[0.3: 0.5]	1.8×10^{-16}	6	0.0%	20.0%	9	0.0%	33.3%	0
urine	X - 12101	2.5[2.0: 3.1]	2.2×10^{-16}	10	0.0%	15.6%	7	0.0%	9.5%	0

continued on next page ...

urine	1-methylxanthine	0.4[0.3: 0.5]	2.2×10^{-16}	12	0.0%	27.3%	2	0.0%	0.0%	1
plasma	docosapentaenoate (n3 DPA; 22:5n3)	0.3[0.3: 0.4]	2.5×10^{-16}	8	0.0%	35.7%	6	0.0%	40.0%	0
urine	X - 17357	0.4[0.3: 0.5]	3.1×10^{-16}	6	0.0%	33.3%	10	0.0%	46.7%	0
urine	N-acetylhistidine	0.3[0.2: 0.4]	3.2×10^{-16}	13	0.0%	14.1%	7	0.0%	4.8%	0
urine	glycolithocholate sulfate	0.3[0.3: 0.4]	3.4×10^{-16}	11	0.0%	5.5%	4	0.0%	33.3%	0
urine	X - 12407	3.9[2.8: 5.4]	4.1×10^{-16}	12	0.0%	6.1%	6	0.0%	0.0%	0
plasma	1-oleoylglycerophosphocholine (18:1)	0.4[0.3: 0.5]	4.4×10^{-16}	6	0.0%	20.0%	12	0.0%	50.0%	0
urine	X - 18345	2.6[2.0: 3.2]	5.7×10^{-16}	5	0.0%	80.0%	6	0.0%	80.0%	0
plasma	uridine	0.4[0.3: 0.5]	7.0×10^{-16}	15	0.0%	1.0%	3	0.0%	33.3%	0
plasma	1-arachidonoylglycerophosphocholine (20:4n6)	0.3[0.3: 0.4]	7.0×10^{-16}	11	0.0%	41.8%	8	0.0%	32.1%	0
plasma	N6-carbamoylthreonyladenosine	109.1[34.9:341.2]	7.2×10^{-16}	8	0.0%	3.6%	0	0.0%	—	0
urine	X - 12093	0.3[0.2: 0.4]	8.1×10^{-16}	11	0.0%	18.2%	9	0.0%	25.0%	0
plasma	17-methylstearate	0.4[0.3: 0.5]	9.0×10^{-16}	4	0.0%	33.3%	6	0.0%	26.7%	0
urine	urate	0.4[0.3: 0.5]	9.1×10^{-16}	10	0.0%	11.1%	1	0.0%	—	0
saliva	urea	4.3[3.0: 6.1]	9.5×10^{-16}	9	0.0%	13.9%	4	0.0%	66.7%	0
plasma	X - 11795	0.3[0.2: 0.4]	9.7×10^{-16}	11	0.0%	0.0%	2	0.0%	0.0%	0
saliva	4-methyl-2-oxopentanoate	0.4[0.3: 0.5]	1.1×10^{-15}	6	0.0%	20.0%	4	0.0%	16.7%	0
plasma	1-oleoylplasmenylethanolamine	0.4[0.3: 0.5]	1.1×10^{-15}	15	0.0%	1.9%	12	0.0%	0.0%	0
urine	X - 12267	2.8[2.2: 3.7]	1.2×10^{-15}	14	0.0%	12.1%	3	0.0%	0.0%	0
plasma	X - 09789	3.0[2.3: 4.0]	1.3×10^{-15}	18	0.0%	3.9%	1	0.0%	—	0
saliva	X - 21365	3.0[2.3: 3.9]	1.4×10^{-15}	16	0.0%	12.5%	11	0.0%	5.5%	0
urine	X - 15863	2.7[2.1: 3.4]	1.8×10^{-15}	11	0.0%	10.9%	0	0.0%	—	0
plasma	X - 16394	99.6[32.1:309.6]	1.8×10^{-15}	13	0.0%	3.8%	2	0.0%	100.0%	0
plasma	1-stearoylglycerophosphocholine (18:0)	0.3[0.3: 0.5]	1.8×10^{-15}	5	0.0%	20.0%	7	0.0%	47.6%	0
saliva	citrate	0.3[0.3: 0.4]	2.3×10^{-15}	9	0.0%	13.9%	10	0.0%	11.1%	0
urine	phenylacetylglutamine	3.6[2.6: 5.0]	2.5×10^{-15}	3	0.0%	100.0%	3	0.0%	33.3%	0
urine	X - 12124	0.3[0.2: 0.4]	3.0×10^{-15}	13	0.0%	25.6%	3	0.0%	100.0%	0
plasma	creatinine	121.6[36.8:401.4]	3.3×10^{-15}	14	0.0%	1.1%	0	0.0%	—	0
urine	N2,N5-diacetylornithine	0.3[0.3: 0.5]	3.3×10^{-15}	7	0.0%	4.8%	5	0.0%	10.0%	0
urine	4-androsten-3beta,17beta-diol disulfate (1)	0.2[0.2: 0.3]	3.5×10^{-15}	11	0.0%	25.5%	7	0.0%	38.1%	0
plasma	3-methoxytyrosine	3.8[2.7: 5.3]	3.6×10^{-15}	12	0.0%	1.5%	7	0.0%	19.0%	0
urine	pyridoxate	2.3[1.9: 2.9]	3.7×10^{-15}	14	0.0%	4.4%	3	0.0%	66.7%	0
urine	pyroglutamine	0.3[0.3: 0.4]	3.9×10^{-15}	8	0.0%	14.3%	1	0.0%	—	0
plasma	1-eicosapentaenoylglycerophosphocholine (20:5n3)	0.3[0.3: 0.4]	3.9×10^{-15}	12	0.0%	24.2%	6	0.0%	33.3%	0
urine	X - 12014	3.3[2.5: 4.5]	4.1×10^{-15}	10	0.0%	11.1%	5	0.0%	10.0%	0
saliva	X - 19841	0.4[0.3: 0.5]	5.0×10^{-15}	7	0.0%	14.3%	9	0.0%	41.7%	0
plasma	pregn steroid monosulfate	0.3[0.2: 0.4]	6.6×10^{-15}	8	0.0%	50.0%	6	0.0%	80.0%	3
urine	X - 12738	2.5[2.0: 3.1]	7.6×10^{-15}	9	0.0%	25.0%	5	0.0%	30.0%	0
urine	X - 12846	0.4[0.3: 0.5]	8.2×10^{-15}	11	0.0%	10.9%	4	0.0%	50.0%	1
urine	X - 12339	0.3[0.3: 0.4]	8.7×10^{-15}	18	0.0%	2.6%	6	0.0%	0.0%	1
plasma	X - 11381	0.4[0.3: 0.5]	1.2×10^{-14}	16	0.0%	5.8%	4	0.0%	0.0%	0
plasma	N1-methyladenosine	2.8[2.1: 3.6]	1.2×10^{-14}	15	0.0%	4.8%	5	0.0%	10.0%	0
urine	kynurenine	2.7[2.1: 3.5]	1.3×10^{-14}	12	0.0%	7.6%	1	0.0%	—	0
plasma	bilirubin (E,E)	0.4[0.3: 0.5]	1.4×10^{-14}	8	0.0%	7.1%	1	0.0%	—	0
plasma	stachydrine	3.2[2.4: 4.4]	1.5×10^{-14}	4	0.0%	83.3%	7	0.0%	28.6%	0
plasma	X - 02249	3.2[2.4: 4.3]	1.7×10^{-14}	7	0.0%	9.5%	1	0.0%	—	0
plasma	isovalerylcarnitine	0.3[0.3: 0.4]	1.8×10^{-14}	16	0.0%	2.5%	7	0.0%	38.1%	0
plasma	Isobar: glucose, fructose, mannose, galactose, allose, altrose, etc.	2.9[2.2: 3.7]	1.8×10^{-14}	11	0.0%	1.8%	13	0.0%	7.7%	2
urine	X - 16581	2.8[2.2: 3.7]	2.0×10^{-14}	15	0.0%	8.6%	10	0.0%	2.2%	0
urine	X - 11334	2.4[1.9: 2.9]	2.2×10^{-14}	11	0.0%	5.5%	2	0.0%	0.0%	0
urine	pantothenate	0.4[0.3: 0.5]	2.2×10^{-14}	8	0.0%	17.9%	7	0.0%	9.5%	0
urine	X - 20634	2.8[2.1: 3.6]	2.3×10^{-14}	10	0.0%	4.4%	3	0.0%	33.3%	0
urine	X - 13722	2.7[2.1: 3.5]	2.4×10^{-14}	8	0.0%	28.6%	5	0.0%	40.0%	0
urine	3-methylglutaconate	2.3[1.9: 2.9]	2.4×10^{-14}	3	0.0%	66.7%	7	0.0%	0.0%	0
urine	X - 18342	2.7[2.1: 3.5]	2.8×10^{-14}	16	0.0%	5.0%	12	0.0%	15.2%	0
urine	5-hydroxyhexanoate	0.3[0.3: 0.4]	2.8×10^{-14}	8	0.0%	3.6%	3	0.0%	33.3%	0
plasma	2-myristoylglycerophosphocholine	0.4[0.3: 0.5]	4.8×10^{-14}	7	0.0%	38.1%	6	0.0%	53.3%	0
saliva	10-hydroxydecanoic acid	2.6[2.0: 3.3]	5.6×10^{-14}	6	0.0%	20.0%	10	0.0%	51.1%	0
urine	xanthine	0.4[0.3: 0.5]	6.2×10^{-14}	9	0.0%	30.6%	7	0.0%	61.9%	1
plasma	4-vinylphenol sulfate	2.6[2.0: 3.3]	6.4×10^{-14}	12	0.0%	10.6%	1	0.0%	—	0
urine	X - 11444	0.4[0.3: 0.5]	6.5×10^{-14}	10	0.0%	35.6%	11	0.0%	34.5%	0
plasma	X - 11787	0.4[0.3: 0.5]	6.6×10^{-14}	8	0.0%	0.0%	3	0.0%	33.3%	0
urine	gentisate	0.4[0.3: 0.5]	6.9×10^{-14}	7	0.0%	14.3%	3	0.0%	0.0%	1
plasma	7-alpha-hydroxy-3-oxo-4-cholestenoate (7-Hoca)	0.4[0.3: 0.5]	8.3×10^{-14}	12	0.0%	4.5%	4	0.0%	0.0%	0
plasma	5-dodecenoate (12:1n7)	0.4[0.3: 0.5]	9.0×10^{-14}	4	0.0%	66.7%	2	0.0%	100.0%	0
plasma	3-(4-hydroxyphenyl)lactate	3.0[2.3: 4.0]	9.0×10^{-14}	15	0.0%	4.8%	1	0.0%	—	0
urine	X - 20588	2.9[2.2: 3.9]	9.6×10^{-14}	4	0.0%	50.0%	4	0.0%	16.7%	1
plasma	X - 12100	196.1[48.8:788.6]	1.0×10^{-13}	10	0.0%	4.4%	0	0.0%	—	0
plasma	1-linolenoylglycerophosphocholine (18:3n3)	0.3[0.3: 0.5]	1.1×10^{-13}	13	0.0%	20.5%	10	0.0%	44.4%	0
plasma	linoleate (18:2n6)	0.4[0.3: 0.5]	1.2×10^{-13}	9	0.0%	13.9%	4	0.0%	50.0%	0
plasma	1-linoleoylglycerophosphocholine (18:2n6)	0.4[0.3: 0.5]	1.2×10^{-13}	11	0.0%	40.0%	11	0.0%	49.1%	0
urine	7-methylxanthine	0.4[0.3: 0.5]	1.4×10^{-13}	7	0.0%	61.9%	7	0.0%	47.6%	2
urine	X - 10593	2.8[2.1: 3.7]	1.5×10^{-13}	15	0.0%	1.0%	4	0.0%	16.7%	0
plasma	1-pentadecanoylglycerophosphocholine (15:0)	0.4[0.3: 0.5]	2.4×10^{-13}	11	0.0%	14.5%	6	0.0%	60.0%	0
urine	X - 11521	0.4[0.3: 0.5]	2.7×10^{-13}	8	0.0%	42.9%	6	0.0%	53.3%	0
saliva	stachydrine	2.6[2.0: 3.4]	3.2×10^{-13}	14	0.0%	11.0%	5	0.0%	60.0%	0
urine	X - 12686	2.9[2.2: 3.8]	3.3×10^{-13}	7	0.0%	14.3%	3	0.0%	0.0%	0
saliva	vanillin	2.7[2.0: 3.5]	3.6×10^{-13}	4	0.0%	33.3%	5	0.0%	80.0%	1
plasma	2-ethylhexanoate (isobar with 2-propylpentanoate)	2.1[1.7: 2.5]	5.2×10^{-13}	19	0.0%	2.3%	9	0.0%	19.4%	0

continued on next page ...

urine	homovanillate (HVA)	0.3[0.2: 0.4]	7.6×10^{-13}	10	0.0%	4.4%	9	0.0%	13.9%	0
urine	X - 12026	3.5[2.5: 5.0]	8.1×10^{-13}	14	0.0%	13.2%	1	0.0%	—	1
saliva	X - 11508	2.5[1.9: 3.2]	8.7×10^{-13}	15	0.0%	4.8%	12	0.0%	15.2%	0
plasma	X - 11727	3.0[2.2: 4.1]	9.1×10^{-13}	8	0.0%	14.3%	6	0.0%	20.0%	2
urine	X - 12511	0.3[0.3: 0.5]	9.4×10^{-13}	9	0.0%	13.9%	7	0.0%	0.0%	0
plasma	1-docosahexaenoylglycerophosphocholine (22:6n3)	0.4[0.3: 0.5]	1.1×10^{-12}	13	0.0%	24.4%	2	0.0%	100.0%	0
plasma	valine	0.4[0.3: 0.5]	1.2×10^{-12}	12	0.0%	9.1%	5	0.0%	40.0%	0
urine	isovalerylcarnitine	0.4[0.3: 0.5]	1.2×10^{-12}	9	0.0%	38.9%	4	0.0%	66.7%	0
urine	succinylcarnitine	0.4[0.3: 0.5]	1.2×10^{-12}	14	0.0%	2.2%	9	0.0%	2.8%	0
saliva	X - 19847	0.5[0.4: 0.6]	1.8×10^{-12}	6	0.0%	53.3%	9	0.0%	41.7%	0
plasma	10-nonadecenoate (19:1n9)	0.4[0.3: 0.5]	1.9×10^{-12}	7	0.0%	38.1%	3	0.0%	66.7%	0
plasma	dihomo-linoleate (20:2n6)	0.4[0.3: 0.5]	2.0×10^{-12}	8	0.0%	35.7%	7	0.0%	33.3%	0
saliva	X - 19846	0.5[0.4: 0.6]	2.0×10^{-12}	8	0.0%	17.9%	6	0.0%	73.3%	0
urine	X - 12306	0.4[0.3: 0.5]	2.1×10^{-12}	17	0.0%	5.9%	7	0.0%	0.0%	0
plasma	myristoleate (14:1n5)	0.4[0.3: 0.5]	2.1×10^{-12}	7	0.0%	52.4%	4	0.0%	33.3%	0
urine	X - 12112	0.4[0.4: 0.6]	2.6×10^{-12}	10	0.0%	13.3%	3	0.0%	33.3%	0
saliva	nicotinate ribonucleoside	0.4[0.3: 0.5]	3.0×10^{-12}	9	0.0%	0.0%	4	0.0%	0.0%	0
plasma	7-methylxanthine	2.5[1.9: 3.2]	3.1×10^{-12}	9	0.0%	2.8%	11	0.0%	16.4%	1
urine	phosphate	3.0[2.2: 4.0]	3.4×10^{-12}	10	0.0%	13.3%	1	0.0%	—	0
urine	X - 14318	2.1[1.7: 2.6]	3.7×10^{-12}	6	0.0%	80.0%	6	0.0%	73.3%	0
urine	X - 12100	2.4[1.9: 3.0]	3.9×10^{-12}	9	0.0%	8.3%	2	0.0%	0.0%	0
urine	X - 14838	2.5[1.9: 3.2]	4.0×10^{-12}	4	0.0%	83.3%	5	0.0%	40.0%	0
urine	X - 12170	2.4[1.9: 3.0]	4.0×10^{-12}	6	0.0%	6.7%	3	0.0%	33.3%	0
plasma	10-heptadecenoate (17:1n7)	0.4[0.3: 0.5]	4.3×10^{-12}	7	0.0%	52.4%	3	0.0%	33.3%	0
plasma	X - 12798	0.4[0.3: 0.6]	4.6×10^{-12}	15	0.0%	4.8%	11	0.0%	7.3%	1
urine	cortisone	0.4[0.3: 0.5]	4.8×10^{-12}	8	0.0%	25.0%	10	0.0%	2.2%	3
saliva	3-phenylpropionate (hydrocinnamate)	0.4[0.3: 0.5]	5.0×10^{-12}	12	0.0%	10.6%	8	0.0%	3.6%	1
plasma	2-aminoheptanoate	2.6[2.0: 3.5]	5.9×10^{-12}	8	0.0%	0.0%	3	0.0%	0.0%	0
urine	X - 17692	2.4[1.9: 3.0]	6.6×10^{-12}	13	0.0%	7.7%	7	0.0%	4.8%	0
urine	X - 14352	2.1[1.7: 2.5]	7.0×10^{-12}	6	0.0%	60.0%	7	0.0%	57.1%	0
saliva	X - 12237	2.5[1.9: 3.3]	7.1×10^{-12}	7	0.0%	19.0%	5	0.0%	60.0%	0
saliva	pantothenate	2.3[1.8: 2.9]	8.1×10^{-12}	11	0.0%	1.8%	7	0.0%	4.8%	0
plasma	palmitoleate (16:1n7)	0.4[0.3: 0.5]	8.2×10^{-12}	8	0.0%	28.6%	4	0.0%	16.7%	0
urine	epiandrosterone sulfate	0.4[0.3: 0.5]	8.4×10^{-12}	9	0.0%	36.1%	7	0.0%	23.8%	0
plasma	gamma-glutamylleucine	2.8[2.1: 3.7]	9.5×10^{-12}	9	0.0%	33.3%	7	0.0%	33.3%	0
urine	3-[3-(sulfooxy)phenyl]propanoic acid	2.7[2.0: 3.6]	1.2×10^{-11}	10	0.0%	17.8%	4	0.0%	83.3%	0
urine	N-acetylglutamine	0.3[0.3: 0.5]	1.2×10^{-11}	11	0.0%	41.8%	6	0.0%	33.3%	0
plasma	X - 12104	483.1[80.82888.5]	1.3×10^{-11}	9	0.0%	5.6%	1	0.0%	—	0
saliva	X - 19869	0.4[0.3: 0.6]	1.7×10^{-11}	6	0.0%	26.7%	10	0.0%	31.1%	0
saliva	imidazole propionate	2.5[1.9: 3.3]	1.9×10^{-11}	11	0.0%	9.1%	10	0.0%	13.3%	0
saliva	X - 19489	2.4[1.9: 3.1]	2.1×10^{-11}	7	0.0%	14.3%	8	0.0%	60.7%	1
plasma	C-glycosyltryptophan	420.2[70.92490.3]	2.9×10^{-11}	8	0.0%	10.7%	0	0.0%	—	0
urine	2-methylbutyrylglycine	2.3[1.8: 2.9]	2.9×10^{-11}	9	0.0%	16.7%	8	0.0%	32.1%	0
urine	histidine	0.4[0.3: 0.5]	3.2×10^{-11}	11	0.0%	9.1%	3	0.0%	33.3%	0
urine	octanoylcarnitine	0.4[0.3: 0.5]	3.2×10^{-11}	8	0.0%	35.7%	7	0.0%	38.1%	0
saliva	X - 12259	2.3[1.8: 2.9]	3.4×10^{-11}	8	0.0%	14.3%	4	0.0%	33.3%	0
urine	dehydroisoandrosterone sulfate (DHEA-S)	0.4[0.3: 0.5]	3.4×10^{-11}	11	0.0%	47.3%	6	0.0%	60.0%	0
plasma	2-palmitoleoylglycerophosphocholine	0.4[0.3: 0.6]	3.6×10^{-11}	11	0.0%	20.0%	7	0.0%	71.4%	0
urine	X - 16394	2.4[1.9: 3.1]	4.0×10^{-11}	10	0.0%	4.4%	6	0.0%	13.3%	1
urine	3-methylglutarate	2.2[1.7: 2.7]	4.7×10^{-11}	9	0.0%	13.9%	5	0.0%	10.0%	1
saliva	X - 15605	0.5[0.4: 0.6]	4.8×10^{-11}	19	0.0%	4.7%	8	0.0%	10.7%	1
plasma	histidine	0.4[0.3: 0.6]	5.0×10^{-11}	17	0.0%	5.1%	7	0.0%	4.8%	0
plasma	oleate (18:1n9)	0.5[0.4: 0.6]	6.5×10^{-11}	8	0.0%	32.1%	7	0.0%	38.1%	0
saliva	X - 19852	0.5[0.4: 0.6]	7.5×10^{-11}	6	0.0%	53.3%	7	0.0%	52.4%	1
urine	N-carbamoylsarcosine	2.5[1.9: 3.3]	7.8×10^{-11}	15	0.0%	1.9%	1	0.0%	—	0
plasma	margarate (17:0)	0.4[0.3: 0.6]	8.9×10^{-11}	7	0.0%	42.9%	2	0.0%	100.0%	0
urine	X - 12122	2.3[1.8: 3.0]	9.4×10^{-11}	14	0.0%	6.6%	2	0.0%	100.0%	0
urine	X - 20643	0.5[0.4: 0.6]	9.7×10^{-11}	14	0.0%	17.6%	9	0.0%	41.7%	0
urine	X - 16087	0.4[0.3: 0.6]	10.0×10^{-11}	7	0.0%	23.8%	1	0.0%	—	0
urine	X - 16975	0.4[0.3: 0.6]	1.0×10^{-10}	6	0.0%	60.0%	2	0.0%	0.0%	0
plasma	glycocholate	2.1[1.7: 2.6]	1.1×10^{-10}	17	0.0%	7.4%	10	0.0%	13.3%	0
urine	X - 17438	0.4[0.3: 0.5]	1.3×10^{-10}	13	0.0%	9.0%	6	0.0%	13.3%	0
saliva	citrulline	0.4[0.3: 0.6]	1.4×10^{-10}	11	0.0%	14.5%	10	0.0%	28.9%	1
saliva	proline	0.4[0.3: 0.6]	1.4×10^{-10}	16	0.0%	7.5%	10	0.0%	17.8%	0
urine	X - 11687	2.1[1.7: 2.7]	1.5×10^{-10}	12	0.0%	1.5%	2	0.0%	100.0%	0
plasma	1-oleoylglycerol (1-monoolein)	2.1[1.7: 2.7]	1.6×10^{-10}	11	0.0%	7.3%	2	0.0%	100.0%	0
plasma	2-linoleoylglycerophosphocholine	0.4[0.3: 0.5]	1.8×10^{-10}	11	0.0%	30.9%	6	0.0%	80.0%	0
urine	gamma-CEHC glucuronide	2.3[1.8: 3.0]	2.0×10^{-10}	7	0.0%	4.8%	4	0.0%	33.3%	0
urine	adipate	2.0[1.6: 2.6]	2.0×10^{-10}	15	0.0%	6.7%	6	0.0%	6.7%	0
plasma	1-docosapentaenoylglycerophosphocholine (22:5n3)	0.5[0.4: 0.6]	2.0×10^{-10}	13	0.0%	20.5%	4	0.0%	33.3%	0
urine	X - 12824	2.4[1.8: 3.1]	2.1×10^{-10}	9	0.0%	5.6%	7	0.0%	23.8%	0
saliva	X - 19572	0.4[0.3: 0.5]	2.1×10^{-10}	6	0.0%	13.3%	5	0.0%	20.0%	0
plasma	2-oleoylglycerophosphocholine	0.4[0.3: 0.6]	2.4×10^{-10}	7	0.0%	42.9%	13	0.0%	44.9%	0
urine	alpha-CEHC glucuronide	2.1[1.7: 2.7]	2.5×10^{-10}	3	0.0%	66.7%	11	0.0%	7.3%	1
saliva	glycerophosphorylcholine (GPC)	2.3[1.8: 2.9]	2.7×10^{-10}	12	0.0%	7.6%	5	0.0%	10.0%	0
urine	N2,N2-dimethylguanosine	2.3[1.8: 3.1]	2.8×10^{-10}	7	0.0%	23.8%	5	0.0%	60.0%	0
urine	X - 18965	0.5[0.4: 0.6]	3.7×10^{-10}	8	0.0%	3.6%	1	0.0%	—	0
saliva	X - 12128	2.3[1.8: 3.0]	3.7×10^{-10}	9	0.0%	19.4%	4	0.0%	50.0%	0
urine	phenylcarnitine	0.4[0.3: 0.6]	4.2×10^{-10}	16	0.0%	5.0%	6	0.0%	6.7%	1

continued on next page ...

urine	theobromine	0.5[0.4: 0.6]	4.8×10^{-10}	7	0.0%	57.1%	7	0.0%	38.1%	1
saliva	N6-acetyllysine	2.1[1.6: 2.6]	4.8×10^{-10}	12	0.0%	10.6%	3	0.0%	0.0%	0
saliva	N-delta-acetylmethionine	2.1[1.7: 2.7]	5.2×10^{-10}	10	0.0%	8.9%	6	0.0%	0.0%	0
plasma	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	3.3[2.2: 4.7]	6.1×10^{-10}	7	0.0%	19.0%	4	0.0%	33.3%	0
urine	X - 20620	2.2[1.7: 2.8]	6.8×10^{-10}	10	0.0%	4.4%	7	0.0%	4.8%	0
plasma	lysine	0.5[0.4: 0.6]	7.0×10^{-10}	12	0.0%	4.5%	9	0.0%	5.6%	0
saliva	X - 14952	0.4[0.3: 0.6]	7.2×10^{-10}	9	0.0%	8.3%	7	0.0%	4.8%	0
plasma	3-dehydrocarnitine	2.3[1.7: 2.9]	7.3×10^{-10}	19	0.0%	6.4%	5	0.0%	10.0%	0
urine	5-acetylamino-6-formylamino-3-methyluracil	0.5[0.4: 0.6]	7.3×10^{-10}	13	0.0%	14.1%	9	0.0%	22.2%	0
plasma	palmitate (16:0)	0.5[0.4: 0.6]	7.9×10^{-10}	6	0.0%	33.3%	4	0.0%	50.0%	0
urine	X - 18886	2.1[1.6: 2.6]	8.0×10^{-10}	11	0.0%	0.0%	7	0.0%	0.0%	0
urine	paraxanthine	0.5[0.4: 0.6]	8.2×10^{-10}	9	0.0%	41.7%	8	0.0%	64.3%	0
urine	X - 12358	2.4[1.8: 3.3]	8.6×10^{-10}	6	0.0%	60.0%	5	0.0%	40.0%	0
urine	X - 12837	0.5[0.4: 0.6]	1.0×10^{-9}	2	0.0%	0.0%	5	0.0%	20.0%	0
plasma	1-eicosenoylglycerophosphocholine (20:1n9)	0.5[0.4: 0.6]	1.2×10^{-9}	11	0.0%	3.6%	4	0.0%	50.0%	0
urine	X - 17308	2.3[1.7: 3.0]	1.5×10^{-9}	7	0.0%	33.3%	4	0.0%	0.0%	0
urine	X - 12258	0.5[0.4: 0.6]	1.5×10^{-9}	7	0.0%	23.8%	4	0.0%	33.3%	1
urine	tigloylglycine	2.1[1.7: 2.7]	1.6×10^{-9}	10	0.0%	22.2%	6	0.0%	40.0%	1
urine	tyramine	0.4[0.3: 0.6]	2.1×10^{-9}	16	0.0%	0.8%	2	0.0%	0.0%	0
plasma	21-hydroxypregnenolone disulfate	2.3[1.7: 3.0]	2.3×10^{-9}	11	0.0%	21.8%	9	0.0%	66.7%	0
saliva	X - 13230	0.5[0.4: 0.7]	2.3×10^{-9}	9	0.0%	19.4%	7	0.0%	19.0%	0
plasma	proline	2.1[1.7: 2.7]	2.4×10^{-9}	15	0.0%	4.8%	6	0.0%	13.3%	0
urine	indolelactate	2.2[1.7: 2.9]	3.0×10^{-9}	12	0.0%	7.6%	3	0.0%	0.0%	0
plasma	15-methylpalmitate (isobar with 2-methylpalmitate)	0.5[0.4: 0.6]	4.1×10^{-9}	9	0.0%	27.8%	3	0.0%	33.3%	0
plasma	adrenate (22:4n6)	0.5[0.4: 0.6]	4.2×10^{-9}	7	0.0%	33.3%	3	0.0%	66.7%	0
urine	X - 18557	2.0[1.6: 2.6]	4.4×10^{-9}	15	0.0%	5.7%	2	0.0%	0.0%	0
urine	X - 12116	2.2[1.7: 2.9]	4.5×10^{-9}	19	0.0%	3.5%	13	0.0%	5.1%	0
urine	N-acetylvaline	0.4[0.3: 0.6]	4.9×10^{-9}	15	0.0%	1.9%	5	0.0%	30.0%	0
urine	X - 16774	2.2[1.7: 2.9]	6.8×10^{-9}	17	0.0%	5.1%	8	0.0%	14.3%	0
saliva	3-methyl-2-oxovalerate	0.5[0.4: 0.6]	7.6×10^{-9}	5	0.0%	20.0%	6	0.0%	6.7%	0
urine	X - 12860	0.5[0.4: 0.6]	7.8×10^{-9}	9	0.0%	22.2%	0	0.0%	—	0
saliva	betaine	2.0[1.6: 2.6]	8.4×10^{-9}	16	0.0%	6.7%	4	0.0%	33.3%	0
urine	N-acetylleucine	2.0[1.6: 2.6]	9.5×10^{-9}	5	0.0%	0.0%	4	0.0%	16.7%	0
urine	X - 17695	1.9[1.5: 2.3]	1.2×10^{-8}	10	0.0%	4.4%	0	0.0%	—	0
urine	X - 12745	0.5[0.4: 0.7]	1.2×10^{-8}	10	0.0%	13.3%	2	0.0%	0.0%	0
plasma	linolenate [alpha or gamma; (18:3n3 or 6)]	0.5[0.4: 0.6]	1.3×10^{-8}	7	0.0%	23.8%	4	0.0%	33.3%	0
saliva	levulinate (4-oxovalerate)	2.0[1.6: 2.6]	1.5×10^{-8}	7	0.0%	19.0%	6	0.0%	26.7%	0
plasma	X - 13435	2.2[1.7: 2.9]	1.5×10^{-8}	9	0.0%	11.1%	7	0.0%	42.9%	0
urine	3-hydroxysebacate	2.0[1.6: 2.5]	1.8×10^{-8}	15	0.0%	6.7%	6	0.0%	13.3%	0
urine	pro-hydroxy-pro	0.5[0.4: 0.6]	2.1×10^{-8}	7	0.0%	19.0%	6	0.0%	20.0%	0
plasma	caprylate (8:0)	1.8[1.4: 2.1]	2.3×10^{-8}	11	0.0%	0.0%	2	0.0%	0.0%	0
urine	X - 12718	2.2[1.7: 3.0]	2.5×10^{-8}	10	0.0%	11.1%	5	0.0%	20.0%	0
plasma	X - 11529	1.9[1.5: 2.4]	2.6×10^{-8}	12	0.0%	16.7%	7	0.0%	9.5%	1
plasma	nonadecanoate (19:0)	0.5[0.4: 0.6]	2.7×10^{-8}	9	0.0%	16.7%	7	0.0%	42.9%	0
urine	gamma-glutamylthreonine	0.5[0.4: 0.6]	2.8×10^{-8}	9	0.0%	27.8%	9	0.0%	33.3%	0
urine	X - 12636	2.2[1.7: 2.9]	2.9×10^{-8}	5	0.0%	40.0%	5	0.0%	50.0%	0
urine	acetylcarnitine	0.5[0.4: 0.7]	3.0×10^{-8}	6	0.0%	86.7%	6	0.0%	60.0%	0
plasma	X - 11440	2.1[1.6: 2.8]	3.1×10^{-8}	5	0.0%	30.0%	9	0.0%	41.7%	0
saliva	Isobar: glucose, fructose, mannose, galactose, allose, altrose, etc.	1.7[1.4: 2.1]	3.1×10^{-8}	6	0.0%	26.7%	7	0.0%	38.1%	0
urine	X - 12740	0.5[0.4: 0.7]	3.1×10^{-8}	9	0.0%	2.8%	8	0.0%	3.6%	0
urine	X - 17736	0.5[0.4: 0.6]	5.0×10^{-8}	15	0.0%	0.0%	10	0.0%	15.6%	2
urine	X - 13452	2.0[1.6: 2.6]	6.4×10^{-8}	13	0.0%	6.4%	10	0.0%	4.4%	0
urine	hippurate	2.1[1.6: 2.8]	7.8×10^{-8}	7	0.0%	42.9%	5	0.0%	40.0%	0
urine	X - 15678	2.0[1.6: 2.6]	7.8×10^{-8}	8	0.0%	3.6%	1	0.0%	—	0
plasma	eicosenoate (20:1n9 or 11)	0.5[0.4: 0.7]	8.4×10^{-8}	6	0.0%	33.3%	7	0.0%	38.1%	0
plasma	1-palmitoylglycerophosphate	0.5[0.4: 0.7]	8.9×10^{-8}	22	0.0%	3.5%	6	0.0%	0.0%	0
plasma	gamma-glutamylmethionine	0.5[0.4: 0.7]	9.2×10^{-8}	13	0.0%	2.6%	4	0.0%	0.0%	0
urine	X - 12123	1.8[1.4: 2.2]	9.8×10^{-8}	5	0.0%	30.0%	5	0.0%	0.0%	1
saliva	prolylglycine	0.5[0.4: 0.7]	1.1×10^{-7}	10	0.0%	4.4%	10	0.0%	8.9%	0
urine	cis-aconitate	0.5[0.4: 0.7]	1.1×10^{-7}	10	0.0%	2.2%	12	0.0%	6.1%	3
saliva	X - 19496	1.9[1.5: 2.4]	1.3×10^{-7}	15	0.0%	9.5%	10	0.0%	57.8%	1
urine	N-acetylputrescine	1.9[1.5: 2.4]	1.3×10^{-7}	8	0.0%	7.1%	5	0.0%	0.0%	0
plasma	leucine	0.5[0.4: 0.6]	1.4×10^{-7}	7	0.0%	14.3%	5	0.0%	50.0%	0
urine	X - 16674	1.9[1.5: 2.4]	1.5×10^{-7}	10	0.0%	8.9%	6	0.0%	13.3%	0
plasma	homostachydrine	1.9[1.5: 2.5]	1.5×10^{-7}	7	0.0%	4.8%	7	0.0%	4.8%	0
plasma	X - 11470	1.9[1.5: 2.5]	1.5×10^{-7}	12	0.0%	15.2%	1	0.0%	—	0
plasma	hydroxybutyrylcarnitine	1.9[1.5: 2.3]	1.5×10^{-7}	9	0.0%	11.1%	5	0.0%	60.0%	0
urine	homostachydrine	0.5[0.4: 0.7]	1.6×10^{-7}	8	0.0%	10.7%	4	0.0%	0.0%	0
plasma	trans-4-hydroxyproline	1.9[1.5: 2.4]	1.6×10^{-7}	14	0.0%	2.2%	10	0.0%	6.7%	0
urine	1,3,7-trimethylurate	2.1[1.6: 2.7]	1.8×10^{-7}	6	0.0%	66.7%	8	0.0%	71.4%	0
urine	X - 17353	1.9[1.5: 2.4]	2.0×10^{-7}	10	0.0%	2.2%	6	0.0%	6.7%	0
urine	isocitrate	0.5[0.4: 0.6]	2.0×10^{-7}	7	0.0%	14.3%	6	0.0%	0.0%	0
saliva	phosphoethanolamine	1.9[1.5: 2.4]	2.1×10^{-7}	10	0.0%	20.0%	8	0.0%	25.0%	0
urine	alpha-hydroxyisovalerate	0.5[0.4: 0.7]	2.1×10^{-7}	13	0.0%	6.4%	4	0.0%	33.3%	1
plasma	beta-hydroxyisovalerate	2.0[1.5: 2.6]	2.3×10^{-7}	8	0.0%	10.7%	2	0.0%	0.0%	0
saliva	12-HETE	2.0[1.5: 2.5]	2.3×10^{-7}	11	0.0%	5.5%	7	0.0%	19.0%	0
urine	N-acetyl-beta-alanine	0.5[0.4: 0.7]	2.3×10^{-7}	11	0.0%	7.3%	5	0.0%	10.0%	0
urine	X - 11357	1.9[1.5: 2.4]	2.4×10^{-7}	10	0.0%	13.3%	8	0.0%	14.3%	0
urine	X - 18256	2.0[1.5: 2.5]	2.6×10^{-7}	8	0.0%	35.7%	4	0.0%	50.0%	0

continued on next page ...

saliva	X - 19845	0.6[0.5: 0.7]	2.8×10^{-7}	13	0.0%	14.1%	9	0.0%	25.0%	0
saliva	4-hydroxybutyrate (GHB)	1.9[1.5: 2.4]	3.0×10^{-7}	7	0.0%	42.9%	9	0.0%	30.6%	1
urine	X - 12822	0.5[0.4: 0.7]	3.2×10^{-7}	9	0.0%	2.8%	0	0.0%	—	0
plasma	xanthine	1.9[1.5: 2.4]	3.8×10^{-7}	12	0.0%	9.1%	7	0.0%	52.4%	0
urine	X - 19807	0.6[0.5: 0.7]	4.1×10^{-7}	9	0.0%	13.9%	3	0.0%	0.0%	0
urine	X - 12212	0.5[0.4: 0.7]	4.6×10^{-7}	9	0.0%	8.3%	8	0.0%	0.0%	0
plasma	1-palmitoylglycerophosphoinositol	0.5[0.4: 0.7]	4.6×10^{-7}	14	0.0%	5.5%	5	0.0%	0.0%	1
urine	X - 17612	0.6[0.4: 0.7]	5.4×10^{-7}	9	0.0%	2.8%	5	0.0%	10.0%	0
saliva	cytidine	0.6[0.4: 0.7]	5.5×10^{-7}	15	0.0%	9.5%	8	0.0%	10.7%	0
plasma	5alpha-pregnan-3beta,20alpha-diol disulfate	1.7[1.4: 2.2]	6.1×10^{-7}	9	0.0%	11.1%	9	0.0%	30.6%	0
saliva	lactate	1.9[1.5: 2.4]	6.4×10^{-7}	12	0.0%	22.7%	9	0.0%	38.9%	0
plasma	X - 12544	0.5[0.4: 0.6]	6.4×10^{-7}	14	0.0%	0.0%	14	0.0%	2.2%	0
plasma	glutamine	0.6[0.4: 0.7]	6.5×10^{-7}	15	0.0%	5.7%	6	0.0%	6.7%	1
saliva	mevalonate	1.9[1.5: 2.5]	7.0×10^{-7}	6	0.0%	40.0%	0	0.0%	—	0
urine	trigonelline (N'-methylnicotinate)	2.0[1.5: 2.7]	7.1×10^{-7}	4	0.0%	33.3%	6	0.0%	33.3%	0
urine	caffeine	2.0[1.5: 2.6]	7.2×10^{-7}	6	0.0%	46.7%	10	0.0%	57.8%	0
urine	X - 12127	0.5[0.4: 0.7]	7.9×10^{-7}	16	0.0%	1.7%	2	0.0%	0.0%	0
urine	X - 18935	2.0[1.5: 2.6]	8.8×10^{-7}	4	0.0%	66.7%	4	0.0%	66.7%	0
saliva	X - 11476	1.8[1.4: 2.3]	9.6×10^{-7}	9	0.0%	2.8%	9	0.0%	41.7%	0
urine	X - 17343	1.9[1.5: 2.5]	9.8×10^{-7}	12	0.0%	12.1%	7	0.0%	19.0%	0
plasma	piperine	0.6[0.4: 0.7]	10.0×10^{-7}	7	0.0%	33.3%	8	0.0%	50.0%	0
urine	N-acetylcarnosine	0.5[0.4: 0.6]	1.2×10^{-6}	11	0.0%	1.8%	5	0.0%	0.0%	0
saliva	choline phosphate	2.3[1.6: 3.1]	1.2×10^{-6}	11	0.0%	10.9%	11	0.0%	9.1%	0
plasma	1-oleoylglycerophosphoinositol	0.6[0.4: 0.7]	1.3×10^{-6}	9	0.0%	8.3%	7	0.0%	4.8%	0
plasma	3-hydroxydecanoate	0.6[0.4: 0.7]	1.4×10^{-6}	7	0.0%	0.0%	2	0.0%	0.0%	0
urine	X - 18927	1.8[1.4: 2.2]	1.4×10^{-6}	20	0.0%	2.6%	4	0.0%	16.7%	0
urine	X - 12816	0.6[0.5: 0.7]	1.4×10^{-6}	1	0.0%	—	4	0.0%	16.7%	0
saliva	maltotetraose	1.8[1.4: 2.3]	1.4×10^{-6}	7	0.0%	19.0%	7	0.0%	14.3%	0
saliva	butyrylcarnitine	1.8[1.4: 2.3]	1.5×10^{-6}	11	0.0%	21.8%	7	0.0%	14.3%	0
saliva	N-acetylphenylalanine	1.7[1.4: 2.2]	1.5×10^{-6}	16	0.0%	6.7%	16	0.0%	5.8%	0
plasma	azelate (nonanedioate)	1.9[1.5: 2.4]	1.6×10^{-6}	14	0.0%	8.8%	7	0.0%	4.8%	1
saliva	X - 12803	1.8[1.4: 2.3]	1.6×10^{-6}	9	0.0%	0.0%	9	0.0%	16.7%	1
urine	N-acetyl-aspartyl-glutamate (NAAG)	0.5[0.4: 0.7]	1.7×10^{-6}	10	0.0%	6.7%	9	0.0%	5.6%	0
urine	X - 18943	1.9[1.5: 2.5]	1.9×10^{-6}	17	0.0%	2.2%	3	0.0%	0.0%	1
saliva	threonylphenylalanine	0.6[0.5: 0.7]	2.0×10^{-6}	6	0.0%	20.0%	5	0.0%	20.0%	0
urine	X - 12637	1.9[1.5: 2.5]	2.0×10^{-6}	6	0.0%	26.7%	5	0.0%	50.0%	0
saliva	fucose	0.6[0.5: 0.7]	2.1×10^{-6}	5	0.0%	20.0%	8	0.0%	7.1%	0
urine	3-hydroxyhippurate	2.0[1.5: 2.6]	2.2×10^{-6}	5	0.0%	80.0%	6	0.0%	46.7%	0
urine	4-guanidinobutanoate	0.5[0.4: 0.7]	2.3×10^{-6}	14	0.0%	4.4%	5	0.0%	20.0%	0
plasma	indoleacetate	1.7[1.3: 2.1]	2.4×10^{-6}	10	0.0%	17.8%	8	0.0%	21.4%	0
plasma	myristate (14:0)	0.6[0.4: 0.7]	2.5×10^{-6}	8	0.0%	42.9%	4	0.0%	33.3%	0
saliva	creatine	0.5[0.4: 0.7]	3.3×10^{-6}	17	0.0%	4.4%	8	0.0%	3.6%	0
plasma	X - 17629	0.3[0.1: 0.5]	3.5×10^{-6}	3	0.0%	33.3%	8	0.0%	35.7%	2
plasma	taurothiocholate 3-sulfate	1.8[1.4: 2.3]	3.5×10^{-6}	10	0.0%	15.6%	10	0.0%	2.2%	0
urine	X - 18603	1.8[1.4: 2.4]	3.6×10^{-6}	8	0.0%	28.6%	6	0.0%	20.0%	0
plasma	2-aminobutyrate	0.6[0.4: 0.7]	3.7×10^{-6}	13	0.0%	2.6%	7	0.0%	0.0%	0
urine	X - 16580	1.8[1.4: 2.3]	3.7×10^{-6}	9	0.0%	5.6%	1	0.0%	—	0
plasma	phenylalanine	1.8[1.4: 2.3]	4.2×10^{-6}	15	0.0%	5.7%	6	0.0%	13.3%	0
urine	pivaloylcarnitine	1.5[1.3: 1.8]	4.2×10^{-6}	8	0.0%	0.0%	9	0.0%	2.8%	0
saliva	caprylate (8:0)	1.8[1.4: 2.4]	4.9×10^{-6}	5	0.0%	50.0%	5	0.0%	60.0%	0
plasma	docosadienoate (22:2n6)	0.6[0.5: 0.7]	6.4×10^{-6}	9	0.0%	8.3%	5	0.0%	70.0%	0
urine	guanine	0.6[0.4: 0.7]	6.9×10^{-6}	11	0.0%	9.1%	4	0.0%	0.0%	0
plasma	alpha-hydroxyisovalerate	0.5[0.4: 0.7]	7.3×10^{-6}	14	0.0%	6.6%	3	0.0%	0.0%	0
urine	phenylalanine	0.6[0.5: 0.7]	7.8×10^{-6}	9	0.0%	13.9%	7	0.0%	71.4%	0
urine	X - 12126	1.9[1.4: 2.6]	8.3×10^{-6}	7	0.0%	19.0%	4	0.0%	66.7%	0
urine	X - 17366	1.7[1.3: 2.1]	9.0×10^{-6}	9	0.0%	5.6%	7	0.0%	4.8%	0
saliva	1,5-anhydroglucitol (1,5-AG)	0.6[0.5: 0.7]	9.5×10^{-6}	6	0.0%	6.7%	5	0.0%	20.0%	0
urine	leucine	1.7[1.3: 2.1]	9.8×10^{-6}	7	0.0%	38.1%	6	0.0%	86.7%	0
urine	7-methylurate	1.9[1.4: 2.6]	9.9×10^{-6}	4	0.0%	16.7%	4	0.0%	0.0%	0
urine	X - 17361	0.6[0.4: 0.7]	1.1×10^{-5}	8	0.0%	7.1%	15	0.0%	4.8%	0
saliva	trans-urocanate	0.6[0.4: 0.7]	1.2×10^{-5}	9	0.0%	11.1%	9	0.0%	8.3%	3
saliva	2,3-dihydroxyisovalerate	1.6[1.3: 2.0]	1.2×10^{-5}	7	0.0%	9.5%	3	0.0%	33.3%	0
plasma	cis-4-decenoyl carnitine	1.8[1.4: 2.3]	1.3×10^{-5}	7	0.0%	23.8%	7	0.0%	47.6%	0
urine	isobutyrylglycine	1.8[1.4: 2.3]	1.3×10^{-5}	8	0.0%	28.6%	4	0.0%	100.0%	0
urine	2-hydroxyhippurate (salicylurate)	1.6[1.3: 1.9]	1.4×10^{-5}	20	0.0%	3.7%	4	0.0%	0.0%	0
saliva	N-acetylneuraminate	0.6[0.5: 0.7]	1.4×10^{-5}	9	0.0%	5.6%	5	0.0%	10.0%	0
urine	3-hydroxyphenylacetate	0.6[0.5: 0.7]	1.4×10^{-5}	11	0.0%	7.3%	13	0.0%	6.4%	2
plasma	methionine	0.6[0.5: 0.7]	1.5×10^{-5}	14	0.0%	7.7%	6	0.0%	20.0%	0
saliva	glutarate (pentanedioate)	1.7[1.3: 2.1]	1.5×10^{-5}	8	0.0%	17.9%	7	0.0%	33.3%	0
urine	carnitine	0.6[0.5: 0.7]	1.6×10^{-5}	9	0.0%	38.9%	5	0.0%	80.0%	0
urine	X - 11835	1.7[1.3: 2.2]	1.6×10^{-5}	8	0.0%	0.0%	8	0.0%	0.0%	0
urine	xanthosine	0.6[0.5: 0.8]	1.7×10^{-5}	11	0.0%	18.2%	9	0.0%	41.7%	0
plasma	androsterone sulfate	0.6[0.5: 0.8]	1.8×10^{-5}	8	0.0%	50.0%	6	0.0%	33.3%	1
urine	N-acetylthreonine	0.6[0.5: 0.8]	1.8×10^{-5}	15	0.0%	11.4%	0	0.0%	—	0
urine	3-methylxanthine	0.6[0.5: 0.8]	1.9×10^{-5}	7	0.0%	47.6%	6	0.0%	66.7%	0
plasma	X - 11491	1.7[1.3: 2.3]	2.2×10^{-5}	14	0.0%	12.1%	10	0.0%	2.2%	0
urine	X - 12832	0.6[0.5: 0.8]	2.3×10^{-5}	11	0.0%	30.9%	7	0.0%	28.6%	0
urine	X - 17354	0.6[0.5: 0.8]	2.6×10^{-5}	10	0.0%	6.7%	5	0.0%	0.0%	0
saliva	4-hydroxyphenylpyruvate	0.6[0.5: 0.8]	2.6×10^{-5}	13	0.0%	1.3%	15	0.0%	7.6%	0

continued on next page ...

plasma	X - 17628	0.3[0.2: 0.5]	2.8×10^{-5}	5	0.0%	10.0%	6	0.0%	60.0%	1
saliva	pipecolate	1.6[1.3: 2.1]	2.8×10^{-5}	14	0.0%	4.4%	7	0.0%	14.3%	0
plasma	cyclo(leu-pro)	1.6[1.3: 2.1]	3.1×10^{-5}	7	0.0%	14.3%	7	0.0%	42.9%	0
plasma	X - 18921	0.6[0.5: 0.8]	3.1×10^{-5}	11	0.0%	12.7%	6	0.0%	13.3%	0
urine	3-methyl-2-oxovalerate	0.6[0.5: 0.8]	3.6×10^{-5}	11	0.0%	10.9%	2	0.0%	0.0%	0
saliva	X - 19848	0.6[0.5: 0.8]	3.8×10^{-5}	12	0.0%	10.6%	6	0.0%	40.0%	0
saliva	sucrose	1.5[1.2: 1.8]	4.0×10^{-5}	8	0.0%	10.7%	8	0.0%	39.3%	0
plasma	alanine	0.6[0.5: 0.8]	4.3×10^{-5}	14	0.0%	6.6%	8	0.0%	10.7%	0
urine	X - 17739	1.6[1.3: 2.0]	5.3×10^{-5}	15	0.0%	6.7%	20	0.0%	2.6%	1
saliva	lysine	0.6[0.5: 0.8]	5.9×10^{-5}	8	0.0%	10.7%	14	0.0%	16.5%	1
plasma	lactate	1.6[1.3: 2.0]	6.7×10^{-5}	12	0.0%	4.5%	9	0.0%	8.3%	1
saliva	X - 18330	0.6[0.5: 0.8]	7.4×10^{-5}	15	0.0%	4.8%	5	0.0%	0.0%	0
plasma	propionylcarnitine	0.6[0.5: 0.8]	7.5×10^{-5}	15	0.0%	7.6%	4	0.0%	100.0%	0
saliva	choline	1.8[1.3: 2.3]	7.7×10^{-5}	12	0.0%	3.0%	10	0.0%	6.7%	0
saliva	X - 15907	0.6[0.5: 0.8]	7.9×10^{-5}	15	0.0%	7.6%	13	0.0%	14.1%	0
saliva	gamma-glutamylleucine	0.6[0.5: 0.8]	8.9×10^{-5}	18	0.0%	6.5%	10	0.0%	4.4%	0
urine	X - 12111	1.6[1.3: 2.1]	9.8×10^{-5}	13	0.0%	12.8%	8	0.0%	21.4%	0
urine	X - 20575	1.6[1.3: 2.0]	1.0×10^{-4}	13	0.0%	3.8%	6	0.0%	20.0%	0
urine	glutaryl carnitine (C5)	0.6[0.5: 0.8]	1.1×10^{-4}	6	0.0%	6.7%	6	0.0%	0.0%	0
saliva	carnitine	0.6[0.5: 0.8]	1.2×10^{-4}	18	0.0%	8.5%	6	0.0%	13.3%	0
saliva	caprate (10:0)	1.7[1.3: 2.2]	1.4×10^{-4}	6	0.0%	13.3%	10	0.0%	26.7%	0
plasma	1-palmitoylglycerol (1-monopalmitin)	1.5[1.2: 1.9]	1.4×10^{-4}	4	0.0%	16.7%	5	0.0%	20.0%	0
saliva	caproate (6:0)	1.7[1.3: 2.1]	1.5×10^{-4}	6	0.0%	33.3%	7	0.0%	38.1%	1
plasma	pregnen-diol disulfate	1.7[1.3: 2.3]	1.6×10^{-4}	10	0.0%	33.3%	9	0.0%	55.6%	0
saliva	urate	0.7[0.5: 0.8]	1.7×10^{-4}	10	0.0%	22.2%	10	0.0%	17.8%	0
plasma	X - 16206	1.6[1.2: 2.0]	1.7×10^{-4}	9	0.0%	19.4%	11	0.0%	10.9%	0
plasma	1-oleoylglycerophosphoethanolamine	1.6[1.2: 2.0]	1.8×10^{-4}	6	0.0%	6.7%	9	0.0%	38.9%	0
saliva	X - 19836	0.6[0.5: 0.8]	1.8×10^{-4}	6	0.0%	20.0%	3	0.0%	0.0%	0
plasma	1-arachidonoylglycerophosphoinositol	0.6[0.5: 0.8]	1.9×10^{-4}	9	0.0%	8.3%	5	0.0%	0.0%	0
saliva	caffeine	1.6[1.2: 2.0]	2.3×10^{-4}	7	0.0%	19.0%	6	0.0%	13.3%	1
plasma	phenylalanintryptophan	1.6[1.2: 2.0]	2.3×10^{-4}	14	0.0%	1.1%	3	0.0%	33.3%	0
urine	succinate	1.5[1.2: 1.8]	2.3×10^{-4}	15	0.0%	7.6%	6	0.0%	13.3%	0
urine	X - 16397	0.6[0.5: 0.8]	2.7×10^{-4}	11	0.0%	3.6%	1	0.0%	—	0
plasma	laurate (12:0)	0.6[0.5: 0.8]	2.8×10^{-4}	9	0.0%	22.2%	4	0.0%	33.3%	0
urine	4-hydroxyphenylacetate	1.5[1.2: 1.8]	2.8×10^{-4}	6	0.0%	13.3%	8	0.0%	3.6%	0
urine	N2-acetyllysine	0.6[0.5: 0.8]	2.9×10^{-4}	11	0.0%	38.2%	9	0.0%	25.0%	0
urine	isoleucine	1.5[1.2: 1.9]	3.1×10^{-4}	6	0.0%	46.7%	7	0.0%	61.9%	0
plasma	EDTA	0.3[0.1: 0.5]	3.3×10^{-4}	6	0.0%	6.7%	6	0.0%	46.7%	4
plasma	X - 11469	0.7[0.5: 0.8]	3.3×10^{-4}	2	0.0%	100.0%	5	0.0%	40.0%	0
saliva	isoleucylglycine	0.6[0.5: 0.8]	3.3×10^{-4}	8	0.0%	21.4%	12	0.0%	6.1%	0
urine	phenylglyoxylic acid	1.5[1.2: 1.8]	4.4×10^{-4}	13	0.0%	1.3%	12	0.0%	4.5%	0
saliva	propionylcarnitine	1.6[1.2: 2.0]	4.6×10^{-4}	5	0.0%	70.0%	3	0.0%	66.7%	0
plasma	X - 02269	0.7[0.5: 0.8]	4.6×10^{-4}	4	0.0%	50.0%	6	0.0%	26.7%	0
plasma	X - 16938	1.6[1.2: 2.0]	5.3×10^{-4}	17	0.0%	5.1%	15	0.0%	1.0%	0
urine	L-urobilin	0.7[0.5: 0.8]	6.6×10^{-4}	18	0.0%	2.6%	11	0.0%	1.8%	0
urine	X - 20502	1.5[1.2: 1.9]	6.8×10^{-4}	12	0.0%	6.1%	2	0.0%	0.0%	0
urine	X - 14302	1.5[1.2: 1.9]	8.0×10^{-4}	11	0.0%	10.9%	3	0.0%	0.0%	0
urine	4-vinylphenol sulfate	0.6[0.5: 0.8]	8.2×10^{-4}	9	0.0%	11.1%	8	0.0%	3.6%	1
urine	X - 13462	1.5[1.2: 1.9]	8.8×10^{-4}	18	0.0%	2.6%	7	0.0%	14.3%	1
urine	3-indoxyl sulfate	1.6[1.2: 2.2]	9.4×10^{-4}	6	0.0%	33.3%	7	0.0%	19.0%	0
plasma	pregnenolone sulfate	0.6[0.4: 0.8]	1.1×10^{-3}	12	0.0%	22.7%	6	0.0%	80.0%	0
saliva	tyrosine	0.7[0.5: 0.9]	1.1×10^{-3}	8	0.0%	17.9%	7	0.0%	23.8%	1
urine	N-methylpipecolate	1.5[1.2: 1.9]	1.1×10^{-3}	3	0.0%	100.0%	8	0.0%	7.1%	0
urine	X - 14951	1.5[1.2: 1.9]	1.2×10^{-3}	15	0.0%	3.8%	13	0.0%	6.4%	1
urine	N-acetylglutamate	1.5[1.2: 1.9]	1.2×10^{-3}	12	0.0%	7.6%	5	0.0%	10.0%	0
urine	gamma-glutamyltyrosine	1.5[1.2: 1.9]	1.2×10^{-3}	9	0.0%	30.6%	5	0.0%	80.0%	0
urine	3-methylcrotonylglycine	1.4[1.2: 1.8]	1.4×10^{-3}	9	0.0%	19.4%	6	0.0%	53.3%	0
urine	theophylline	0.7[0.6: 0.9]	1.4×10^{-3}	8	0.0%	46.4%	8	0.0%	46.4%	0
urine	X - 12386	1.5[1.2: 1.9]	1.5×10^{-3}	11	0.0%	38.2%	13	0.0%	19.2%	0
urine	X - 12689	1.5[1.2: 2.0]	1.6×10^{-3}	14	0.0%	3.3%	5	0.0%	0.0%	0
plasma	S-methylcysteine	0.6[0.5: 0.8]	1.6×10^{-3}	12	0.0%	3.0%	9	0.0%	2.8%	2
urine	cortisol	0.7[0.5: 0.9]	1.6×10^{-3}	9	0.0%	25.0%	8	0.0%	10.7%	1
plasma	stearoylcarnitine	0.7[0.5: 0.9]	1.7×10^{-3}	7	0.0%	14.3%	3	0.0%	33.3%	0
plasma	HWESASXX	0.6[0.5: 0.8]	1.8×10^{-3}	5	0.0%	0.0%	11	0.0%	7.3%	2
saliva	5-oxoproline	1.4[1.1: 1.8]	1.9×10^{-3}	20	0.0%	3.7%	13	0.0%	7.7%	0
urine	4-acetylphenol sulfate	1.5[1.2: 1.9]	1.9×10^{-3}	13	0.0%	6.4%	5	0.0%	20.0%	0
urine	X - 12225	0.7[0.6: 0.9]	1.9×10^{-3}	19	0.0%	4.7%	4	0.0%	0.0%	0
saliva	phenyllactate (PLA)	1.5[1.1: 1.9]	2.1×10^{-3}	8	0.0%	25.0%	4	0.0%	0.0%	0
urine	X - 16563	0.7[0.5: 0.9]	2.2×10^{-3}	7	0.0%	0.0%	8	0.0%	7.1%	1
plasma	cortisol	0.7[0.5: 0.9]	2.2×10^{-3}	11	0.0%	14.5%	9	0.0%	27.8%	2
plasma	4-androsten-3beta,17beta-diol disulfate (2)	1.5[1.2: 2.0]	2.4×10^{-3}	11	0.0%	23.6%	8	0.0%	64.3%	0
urine	4-methylcatechol sulfate	0.7[0.6: 0.9]	2.5×10^{-3}	3	0.0%	66.7%	7	0.0%	28.6%	1
urine	N-acetyl-1-methylhistidine	1.5[1.1: 1.9]	2.9×10^{-3}	13	0.0%	5.1%	4	0.0%	66.7%	0
urine	3-hydroxyanthranilate	1.4[1.1: 1.9]	3.3×10^{-3}	11	0.0%	1.8%	16	0.0%	4.2%	0
plasma	X - 11438	1.5[1.1: 1.9]	3.4×10^{-3}	10	0.0%	4.4%	0	0.0%	—	0
urine	4-methyl-2-oxopentanoate	1.4[1.1: 1.8]	3.5×10^{-3}	15	0.0%	2.9%	6	0.0%	6.7%	0
saliva	X - 12944	1.4[1.1: 1.8]	3.9×10^{-3}	17	0.0%	6.6%	7	0.0%	33.3%	0
urine	3,7-dimethylurate	1.5[1.1: 1.9]	3.9×10^{-3}	9	0.0%	38.9%	10	0.0%	17.8%	2
saliva	histidine	0.7[0.6: 0.9]	4.0×10^{-3}	10	0.0%	11.1%	7	0.0%	14.3%	1

continued on next page ...

plasma	glycoursodeoxycholate	1.4[1.1: 1.8]	4.3×10^{-3}	9	0.0%	2.8%	14	0.0%	3.3%	0
plasma	glycolithocholate sulfate	1.4[1.1: 1.9]	4.6×10^{-3}	7	0.0%	19.0%	10	0.0%	8.9%	0
urine	2-oxindole-3-acetate	1.4[1.1: 1.8]	4.8×10^{-3}	12	0.0%	3.0%	11	0.0%	7.3%	0
saliva	X - 14081	0.7[0.6: 0.9]	5.0×10^{-3}	5	0.0%	80.0%	7	0.0%	47.6%	3
plasma	tryptophan betaine	0.7[0.6: 0.9]	5.1×10^{-3}	12	0.0%	6.1%	15	0.0%	5.7%	0
urine	picolinate	1.4[1.1: 1.8]	5.2×10^{-3}	14	0.0%	2.2%	5	0.0%	0.0%	0
urine	gamma-glutamylphenylalanine	0.7[0.6: 0.9]	5.3×10^{-3}	5	0.0%	60.0%	6	0.0%	73.3%	0
saliva	Lewis X trisaccharide	0.7[0.6: 0.9]	6.0×10^{-3}	11	0.0%	9.1%	15	0.0%	1.9%	0
saliva	X - 11854	0.7[0.6: 0.9]	6.1×10^{-3}	5	0.0%	80.0%	7	0.0%	52.4%	3
urine	5-hydroxymethyl-2-furoic acid	1.4[1.1: 1.8]	6.1×10^{-3}	11	0.0%	14.5%	3	0.0%	0.0%	0
saliva	nicotinate	0.7[0.6: 0.9]	6.4×10^{-3}	12	0.0%	1.5%	10	0.0%	15.6%	0
saliva	gamma-glutamylmethionine	0.7[0.6: 0.9]	6.5×10^{-3}	10	0.0%	13.3%	11	0.0%	1.8%	0
plasma	HWESALLR	0.7[0.5: 0.9]	6.6×10^{-3}	4	0.0%	0.0%	5	0.0%	40.0%	1
plasma	caffeine	1.4[1.1: 1.8]	6.8×10^{-3}	9	0.0%	19.4%	7	0.0%	47.6%	1
urine	X - 13866	0.7[0.6: 0.9]	6.8×10^{-3}	22	0.0%	3.5%	5	0.0%	0.0%	0
plasma	caproate (6:0)	1.4[1.1: 1.7]	7.0×10^{-3}	10	0.0%	4.4%	4	0.0%	0.0%	0
saliva	pelargonate (9:0)	1.4[1.1: 1.8]	7.9×10^{-3}	7	0.0%	19.0%	6	0.0%	53.3%	0
urine	X - 15472	0.7[0.6: 0.9]	8.6×10^{-3}	11	0.0%	3.6%	3	0.0%	0.0%	0
saliva	pyruvate	1.4[1.1: 1.8]	8.8×10^{-3}	9	0.0%	13.9%	10	0.0%	31.1%	0
saliva	docosadioate	1.4[1.1: 1.8]	9.6×10^{-3}	16	0.0%	7.5%	11	0.0%	10.9%	1
urine	ascorbate (Vitamin C)	0.7[0.5: 0.9]	9.9×10^{-3}	12	0.0%	0.0%	5	0.0%	10.0%	0
plasma	2-arachidonoylglycerophosphoethanolamine	1.4[1.1: 1.7]	1.1×10^{-2}	13	0.0%	3.8%	6	0.0%	80.0%	0
saliva	phenylacetate	0.7[0.6: 0.9]	1.1×10^{-2}	8	0.0%	25.0%	4	0.0%	50.0%	0
urine	gamma-glutamylleucine	0.7[0.6: 0.9]	1.2×10^{-2}	5	0.0%	90.0%	7	0.0%	47.6%	0
urine	isovalerylglycine	1.4[1.1: 1.7]	1.2×10^{-2}	9	0.0%	16.7%	7	0.0%	38.1%	0
urine	hexanoylglycine	0.7[0.6: 0.9]	1.3×10^{-2}	8	0.0%	14.3%	7	0.0%	14.3%	0
urine	X - 17313	1.4[1.1: 1.8]	1.4×10^{-2}	5	0.0%	80.0%	10	0.0%	17.8%	2
urine	1,3-dimethylurate	1.4[1.1: 1.9]	1.4×10^{-2}	8	0.0%	46.4%	10	0.0%	35.6%	0
plasma	2-piperidinone	1.3[1.1: 1.7]	1.5×10^{-2}	2	0.0%	100.0%	3	0.0%	33.3%	0
urine	3-methyluracil	0.7[0.6: 0.9]	1.5×10^{-2}	3	0.0%	66.7%	3	0.0%	0.0%	1
urine	1,7-dimethylurate	1.5[1.1: 2.0]	1.6×10^{-2}	11	0.0%	40.0%	8	0.0%	50.0%	1
urine	pyroglutamylglutamine	1.3[1.1: 1.7]	1.7×10^{-2}	17	0.0%	5.9%	5	0.0%	10.0%	0
saliva	phosphate	1.3[1.1: 1.7]	1.9×10^{-2}	17	0.0%	12.5%	11	0.0%	21.8%	0
plasma	DSGEGDFXAEAGGVR	1.3[1.0: 1.6]	2.0×10^{-2}	14	0.0%	5.5%	4	0.0%	66.7%	1
urine	1-methylurate	1.4[1.1: 1.8]	2.0×10^{-2}	9	0.0%	41.7%	8	0.0%	32.1%	1
saliva	N-acetylleucine	1.3[1.0: 1.7]	2.0×10^{-2}	11	0.0%	5.5%	6	0.0%	13.3%	0
saliva	uridine	0.8[0.6: 1.0]	2.0×10^{-2}	11	0.0%	12.7%	8	0.0%	3.6%	0
urine	carnosine	0.7[0.6: 1.0]	2.0×10^{-2}	12	0.0%	1.5%	2	0.0%	0.0%	0
plasma	2-aminooctanoate	0.7[0.6: 1.0]	2.2×10^{-2}	8	0.0%	7.1%	8	0.0%	10.7%	0
urine	X - 18554	0.7[0.6: 1.0]	2.2×10^{-2}	19	0.0%	2.9%	11	0.0%	5.5%	0
urine	X - 20617	0.7[0.6: 1.0]	2.3×10^{-2}	19	0.0%	3.5%	6	0.0%	6.7%	0
plasma	1-stearoylglycerophosphoethanolamine	0.8[0.6: 1.0]	2.5×10^{-2}	8	0.0%	17.9%	6	0.0%	46.7%	0
plasma	X - 01911	1.3[1.0: 1.7]	2.7×10^{-2}	4	0.0%	50.0%	6	0.0%	20.0%	0
urine	3-methyl catechol sulfate (1)	1.3[1.0: 1.7]	2.8×10^{-2}	6	0.0%	13.3%	7	0.0%	19.0%	0
saliva	isoleucine	0.8[0.6: 1.0]	2.8×10^{-2}	6	0.0%	40.0%	8	0.0%	35.7%	4
urine	X - 16564	1.3[1.0: 1.7]	2.9×10^{-2}	5	0.0%	70.0%	6	0.0%	40.0%	0
plasma	N-methyl proline	1.3[1.0: 1.7]	3.0×10^{-2}	4	0.0%	50.0%	14	0.0%	9.9%	1
urine	X - 17305	0.8[0.6: 1.0]	3.0×10^{-2}	6	0.0%	13.3%	1	0.0%	—	0
urine	X - 12128	1.3[1.0: 1.7]	3.1×10^{-2}	11	0.0%	10.9%	4	0.0%	0.0%	0
urine	2-methylmalonyl carnitine	0.8[0.6: 1.0]	3.1×10^{-2}	12	0.0%	3.0%	5	0.0%	10.0%	0
saliva	acetylcarnitine	1.3[1.0: 1.7]	3.3×10^{-2}	10	0.0%	24.4%	12	0.0%	13.6%	0
urine	X - 12812	0.8[0.6: 1.0]	3.3×10^{-2}	14	0.0%	2.2%	8	0.0%	3.6%	0
urine	X - 17325	1.3[1.0: 1.7]	3.7×10^{-2}	5	0.0%	70.0%	6	0.0%	40.0%	0
saliva	X - 14473	1.3[1.0: 1.7]	3.9×10^{-2}	7	0.0%	9.5%	3	0.0%	0.0%	0
urine	2,3-dihydroxyisovalerate	1.3[1.0: 1.6]	4.0×10^{-2}	17	0.0%	2.9%	6	0.0%	13.3%	0
plasma	pipecolate	1.2[1.0: 1.5]	4.1×10^{-2}	15	0.0%	4.8%	3	0.0%	0.0%	0
urine	indoleacetate	0.8[0.6: 1.0]	4.3×10^{-2}	13	0.0%	14.1%	7	0.0%	28.6%	0
urine	X - 12283	0.8[0.6: 1.0]	4.7×10^{-2}	7	0.0%	14.3%	2	0.0%	0.0%	0
urine	X - 17686	1.3[1.0: 1.6]	4.9×10^{-2}	10	0.0%	24.4%	10	0.0%	8.9%	0
saliva	valerate	1.3[1.0: 1.6]	5.3×10^{-2}	4	0.0%	50.0%	13	0.0%	14.1%	3
plasma	1-linoleoylglycerophosphoethanolamine	1.3[1.0: 1.6]	5.5×10^{-2}	14	0.0%	26.4%	6	0.0%	86.7%	0
saliva	leucine	0.8[0.6: 1.0]	5.7×10^{-2}	6	0.0%	33.3%	10	0.0%	20.0%	3
urine	S-(3-hydroxypropyl)mercapturic acid (HPMA)	1.2[1.0: 1.6]	5.9×10^{-2}	12	0.0%	3.0%	0	0.0%	—	2
urine	homocitrulline	1.3[1.0: 1.6]	6.0×10^{-2}	9	0.0%	8.3%	1	0.0%	—	0
plasma	theophylline	1.3[1.0: 1.6]	6.2×10^{-2}	5	0.0%	30.0%	9	0.0%	50.0%	0
saliva	lauryl sulfate	0.8[0.6: 1.0]	6.2×10^{-2}	19	0.0%	3.5%	6	0.0%	0.0%	0
saliva	X - 19843	0.8[0.6: 1.0]	6.8×10^{-2}	15	0.0%	7.6%	6	0.0%	13.3%	0
plasma	pyroglutamine	1.3[1.0: 1.8]	6.9×10^{-2}	12	0.0%	9.1%	2	0.0%	0.0%	2
plasma	laurylcarnitine	1.3[1.0: 1.6]	6.9×10^{-2}	8	0.0%	17.9%	4	0.0%	100.0%	0
saliva	3-(4-hydroxyphenyl)lactate	1.3[1.0: 1.6]	7.0×10^{-2}	11	0.0%	14.5%	5	0.0%	10.0%	0
plasma	3-hydroxyisobutyrate	1.2[1.0: 1.6]	7.2×10^{-2}	4	0.0%	50.0%	3	0.0%	100.0%	0
urine	imidazole propionate	1.2[1.0: 1.5]	7.4×10^{-2}	16	0.0%	4.2%	8	0.0%	3.6%	0
plasma	epiandrosterone sulfate	0.8[0.6: 1.0]	7.4×10^{-2}	8	0.0%	53.6%	7	0.0%	23.8%	0
plasma	theobromine	0.8[0.6: 1.0]	7.5×10^{-2}	8	0.0%	46.4%	9	0.0%	27.8%	1
saliva	dexpanthenol	0.8[0.6: 1.0]	7.5×10^{-2}	16	0.0%	0.8%	16	0.0%	3.3%	1
saliva	3-methyl-2-oxobutyrate	1.2[1.0: 1.6]	7.5×10^{-2}	9	0.0%	5.6%	5	0.0%	10.0%	0
urine	X - 16570	1.3[1.0: 1.6]	7.8×10^{-2}	9	0.0%	30.6%	7	0.0%	28.6%	0
urine	X - 18838	1.3[1.0: 1.7]	7.8×10^{-2}	15	0.0%	5.7%	10	0.0%	2.2%	1
plasma	caprate (10:0)	1.2[1.0: 1.5]	8.1×10^{-2}	8	0.0%	17.9%	3	0.0%	0.0%	0

continued on next page ...

urine	X - 02249	0.8[0.6: 1.0]	8.8×10^{-2}	9	0.0%	19.4%	6	0.0%	6.7%	0
plasma	glutamate	1.2[1.0: 1.6]	9.0×10^{-2}	27	0.0%	6.3%	8	0.0%	17.9%	0
plasma	1-arachidonoylglycerophosphoethanolamine	1.2[1.0: 1.5]	9.1×10^{-2}	12	0.0%	27.3%	7	0.0%	76.2%	0
plasma	X - 11521	1.2[1.0: 1.6]	1.0×10^{-1}	14	0.0%	18.7%	6	0.0%	20.0%	0
urine	X - 16940	0.8[0.6: 1.0]	1.0×10^{-1}	4	0.0%	33.3%	1	0.0%	—	0
saliva	4-hydroxyphenylacetate	1.2[1.0: 1.5]	1.0×10^{-1}	11	0.0%	12.7%	11	0.0%	9.1%	0
saliva	X - 16612	0.8[0.7: 1.0]	1.0×10^{-1}	5	0.0%	60.0%	10	0.0%	22.2%	2
saliva	alpha-hydroxyisocaproate	1.2[1.0: 1.5]	1.0×10^{-1}	16	0.0%	6.7%	7	0.0%	4.8%	0
urine	4-ethylphenylsulfate	0.8[0.6: 1.0]	1.1×10^{-1}	18	0.0%	3.3%	10	0.0%	6.7%	0
urine	inosine	0.8[0.6: 1.0]	1.1×10^{-1}	9	0.0%	8.3%	7	0.0%	0.0%	0
urine	6-oxopiperidine-2-carboxylic acid	0.8[0.6: 1.0]	1.2×10^{-1}	8	0.0%	3.6%	10	0.0%	2.2%	0
saliva	spermidine	0.8[0.7: 1.0]	1.2×10^{-1}	11	0.0%	3.6%	11	0.0%	10.9%	0
plasma	decanoylcarnitine	1.2[0.9: 1.6]	1.2×10^{-1}	5	0.0%	60.0%	5	0.0%	80.0%	0
saliva	X - 13205	0.8[0.7: 1.1]	1.3×10^{-1}	5	0.0%	80.0%	5	0.0%	90.0%	0
plasma	octanoylcarnitine	1.2[0.9: 1.5]	1.3×10^{-1}	5	0.0%	40.0%	6	0.0%	66.7%	0
urine	X - 12687	1.2[0.9: 1.6]	1.3×10^{-1}	10	0.0%	17.8%	8	0.0%	14.3%	2
urine	11-ketoetiocholanolone glucuronide	0.8[0.7: 1.1]	1.3×10^{-1}	11	0.0%	7.3%	2	0.0%	100.0%	0
plasma	acetylcarnitine	1.2[0.9: 1.6]	1.3×10^{-1}	8	0.0%	21.4%	9	0.0%	30.6%	0
plasma	heme	1.2[0.9: 1.5]	1.4×10^{-1}	14	0.0%	13.2%	5	0.0%	0.0%	0
saliva	X - 18140	0.8[0.7: 1.1]	1.4×10^{-1}	6	0.0%	26.7%	5	0.0%	60.0%	0
plasma	5alpha-androstan-3beta,17beta-diol disulfate	1.2[0.9: 1.7]	1.5×10^{-1}	8	0.0%	42.9%	6	0.0%	53.3%	0
urine	X - 12261	1.2[0.9: 1.5]	1.6×10^{-1}	3	0.0%	66.7%	10	0.0%	15.6%	0
plasma	andro steroid monosulfate 2	1.2[0.9: 1.5]	1.6×10^{-1}	10	0.0%	26.7%	9	0.0%	41.7%	1
saliva	X - 19867	0.8[0.7: 1.1]	1.8×10^{-1}	11	0.0%	3.6%	4	0.0%	0.0%	0
plasma	pelargonate (9:0)	1.2[0.9: 1.5]	1.8×10^{-1}	5	0.0%	0.0%	9	0.0%	0.0%	2
plasma	indolepropionate	0.8[0.7: 1.1]	1.8×10^{-1}	8	0.0%	7.1%	2	0.0%	0.0%	0
urine	X - 20567	0.9[0.7: 1.1]	1.8×10^{-1}	16	0.0%	5.0%	8	0.0%	10.7%	0
plasma	butyrylcarnitine	0.8[0.6: 1.1]	1.9×10^{-1}	12	0.0%	3.0%	8	0.0%	35.7%	0
urine	4-hydroxybenzoate	1.2[0.9: 1.5]	1.9×10^{-1}	10	0.0%	4.4%	7	0.0%	4.8%	0
plasma	X - 16044	0.9[0.7: 1.1]	1.9×10^{-1}	6	0.0%	13.3%	9	0.0%	5.6%	0
saliva	valine	0.8[0.7: 1.1]	1.9×10^{-1}	7	0.0%	9.5%	8	0.0%	32.1%	2
saliva	X - 18983	0.9[0.7: 1.1]	2.1×10^{-1}	6	0.0%	40.0%	8	0.0%	21.4%	0
urine	X - 11843	1.2[0.9: 1.5]	2.1×10^{-1}	2	0.0%	100.0%	7	0.0%	33.3%	0
plasma	1-palmitoylglycerophosphoethanolamine	1.2[0.9: 1.5]	2.3×10^{-1}	10	0.0%	15.6%	9	0.0%	25.0%	0
saliva	X - 19807	0.9[0.7: 1.1]	2.3×10^{-1}	6	0.0%	20.0%	12	0.0%	6.1%	2
plasma	2-hydroxydecanoate	0.9[0.7: 1.1]	2.4×10^{-1}	8	0.0%	3.6%	8	0.0%	3.6%	0
plasma	2-linoleoylglycerophosphoethanolamine	1.2[0.9: 1.5]	2.5×10^{-1}	17	0.0%	8.1%	7	0.0%	76.2%	0
urine	X - 12828	1.1[0.9: 1.4]	2.6×10^{-1}	9	0.0%	2.8%	4	0.0%	0.0%	0
urine	X - 12695	1.2[0.9: 1.5]	2.6×10^{-1}	9	0.0%	2.8%	8	0.0%	3.6%	1
plasma	X - 11478	0.9[0.7: 1.1]	2.7×10^{-1}	12	0.0%	16.7%	5	0.0%	20.0%	0
saliva	guanosine	1.1[0.9: 1.5]	2.7×10^{-1}	11	0.0%	5.5%	6	0.0%	20.0%	0
urine	3-methoxytyrosine	1.1[0.9: 1.4]	2.8×10^{-1}	13	0.0%	3.8%	9	0.0%	11.1%	0
plasma	hexanoylcarnitine	1.1[0.9: 1.5]	3.0×10^{-1}	9	0.0%	19.4%	6	0.0%	46.7%	0
plasma	3-hydroxybutyrate (BHBA)	1.1[0.9: 1.4]	3.1×10^{-1}	5	0.0%	40.0%	3	0.0%	100.0%	0
saliva	cortisone	0.9[0.7: 1.1]	3.1×10^{-1}	10	0.0%	2.2%	6	0.0%	26.7%	0
saliva	putrescine	1.1[0.9: 1.5]	3.3×10^{-1}	9	0.0%	2.8%	5	0.0%	10.0%	0
plasma	X - 14588	1.1[0.9: 1.4]	3.4×10^{-1}	5	0.0%	30.0%	4	0.0%	66.7%	0
urine	X - 12097	1.1[0.9: 1.4]	3.4×10^{-1}	5	0.0%	50.0%	3	0.0%	33.3%	0
saliva	2-hydroxyglutarate	0.9[0.7: 1.1]	3.5×10^{-1}	15	0.0%	5.7%	11	0.0%	7.3%	1
urine	hexanoylcarnitine	0.9[0.7: 1.1]	3.5×10^{-1}	8	0.0%	0.0%	3	0.0%	0.0%	0
plasma	4-androsten-3beta,17beta-diol disulfate (1)	1.1[0.9: 1.5]	3.6×10^{-1}	6	0.0%	46.7%	6	0.0%	73.3%	0
plasma	pyruvate	1.1[0.9: 1.4]	3.6×10^{-1}	9	0.0%	19.4%	8	0.0%	3.6%	0
plasma	urate	1.1[0.9: 1.5]	3.6×10^{-1}	12	0.0%	3.0%	3	0.0%	0.0%	0
urine	p-cresol sulfate	1.1[0.9: 1.5]	3.7×10^{-1}	6	0.0%	6.7%	8	0.0%	32.1%	1
urine	cinnamoylglycine	0.9[0.7: 1.1]	3.7×10^{-1}	10	0.0%	13.3%	9	0.0%	5.6%	1
plasma	X - 15668	0.9[0.7: 1.1]	3.7×10^{-1}	6	0.0%	6.7%	9	0.0%	5.6%	0
saliva	5-aminovalerate	0.9[0.7: 1.1]	3.7×10^{-1}	12	0.0%	13.6%	6	0.0%	6.7%	0
urine	X - 12216	1.1[0.9: 1.4]	3.9×10^{-1}	7	0.0%	23.8%	9	0.0%	25.0%	1
urine	N-acetylaspartate (NAA)	0.9[0.7: 1.2]	4.0×10^{-1}	8	0.0%	14.3%	6	0.0%	6.7%	0
urine	X - 17688	0.9[0.7: 1.2]	4.1×10^{-1}	9	0.0%	22.2%	9	0.0%	19.4%	0
saliva	X - 19870	0.9[0.7: 1.1]	4.2×10^{-1}	11	0.0%	16.4%	11	0.0%	18.2%	0
saliva	cis, cis-muconic acid	1.1[0.9: 1.4]	4.2×10^{-1}	6	0.0%	33.3%	5	0.0%	30.0%	0
plasma	X - 11538	1.1[0.9: 1.4]	4.2×10^{-1}	18	0.0%	2.6%	1	0.0%	—	0
urine	riboflavin (Vitamin B2)	1.1[0.9: 1.4]	4.3×10^{-1}	13	0.0%	5.1%	8	0.0%	14.3%	0
urine	stachydrine	1.1[0.9: 1.4]	4.4×10^{-1}	7	0.0%	38.1%	6	0.0%	40.0%	1
urine	X - 12095	0.9[0.7: 1.2]	4.5×10^{-1}	4	0.0%	66.7%	4	0.0%	66.7%	0
saliva	guanine	0.9[0.7: 1.2]	4.6×10^{-1}	8	0.0%	14.3%	4	0.0%	16.7%	0
plasma	1-eicosatrienoylglycerophosphoethanolamine	0.9[0.7: 1.2]	4.8×10^{-1}	16	0.0%	6.7%	8	0.0%	39.3%	0
urine	N1-Methyl-2-pyridone-5-carboxamide	1.1[0.9: 1.4]	4.9×10^{-1}	5	0.0%	40.0%	4	0.0%	50.0%	0
saliva	N-acetylserine	1.1[0.9: 1.4]	4.9×10^{-1}	12	0.0%	4.5%	5	0.0%	10.0%	1
urine	alpha-CEHC sulfate	0.9[0.7: 1.2]	5.2×10^{-1}	6	0.0%	20.0%	5	0.0%	20.0%	1
plasma	paraxanthine	0.9[0.7: 1.2]	5.3×10^{-1}	8	0.0%	35.7%	10	0.0%	44.4%	1
urine	glutamine	0.9[0.7: 1.2]	5.5×10^{-1}	6	0.0%	13.3%	8	0.0%	42.9%	2
urine	methionine	0.9[0.7: 1.2]	5.5×10^{-1}	11	0.0%	16.4%	5	0.0%	90.0%	0
urine	X - 17185	1.1[0.8: 1.4]	5.6×10^{-1}	11	0.0%	10.9%	6	0.0%	0.0%	1
urine	2-piperidinone	0.9[0.7: 1.2]	5.6×10^{-1}	3	0.0%	66.7%	5	0.0%	20.0%	0
urine	valine	0.9[0.7: 1.2]	5.7×10^{-1}	11	0.0%	16.4%	5	0.0%	90.0%	0
urine	indoleacetylglutamine	1.1[0.8: 1.4]	5.7×10^{-1}	15	0.0%	5.7%	8	0.0%	14.3%	0
plasma	X - 19808	1.1[0.8: 1.4]	5.7×10^{-1}	5	0.0%	50.0%	4	0.0%	50.0%	0

continued on next page ...

urine	catechol sulfate	0.9[0.7: 1.2]	5.9×10^{-1}	3	0.0%	33.3%	9	0.0%	2.8%	0
urine	X - 14082	1.1[0.8: 1.4]	6.1×10^{-1}	8	0.0%	7.1%	7	0.0%	9.5%	0
plasma	X - 16071	1.1[0.8: 1.4]	6.2×10^{-1}	14	0.0%	4.4%	3	0.0%	0.0%	0
saliva	theobromine	0.9[0.7: 1.2]	6.2×10^{-1}	12	0.0%	10.6%	7	0.0%	14.3%	0
plasma	X - 12776	0.9[0.7: 1.2]	6.4×10^{-1}	12	0.0%	9.1%	12	0.0%	4.5%	1
saliva	inosine	0.9[0.7: 1.2]	6.4×10^{-1}	9	0.0%	11.1%	5	0.0%	40.0%	0
plasma	13-HODE + 9-HODE	0.9[0.7: 1.2]	6.5×10^{-1}	14	0.0%	3.3%	2	0.0%	100.0%	0
urine	X - 11593	0.9[0.7: 1.2]	6.5×10^{-1}	12	0.0%	4.5%	6	0.0%	6.7%	0
urine	X - 12335	0.9[0.7: 1.2]	6.5×10^{-1}	7	0.0%	19.0%	5	0.0%	60.0%	0
plasma	1-docosahexaenoylglycerophosphoethanolamine	1.1[0.8: 1.4]	6.6×10^{-1}	9	0.0%	38.9%	8	0.0%	35.7%	0
urine	X - 12193	0.9[0.7: 1.2]	6.8×10^{-1}	8	0.0%	7.1%	0	0.0%	—	0
plasma	palmitoylcarnitine	0.9[0.7: 1.2]	6.9×10^{-1}	4	0.0%	66.7%	2	0.0%	100.0%	0
saliva	3-(4-hydroxyphenyl)propionate	1.0[0.8: 1.3]	7.0×10^{-1}	11	0.0%	14.5%	8	0.0%	21.4%	0
urine	X - 12753	1.0[0.8: 1.3]	7.1×10^{-1}	17	0.0%	4.4%	13	0.0%	5.1%	0
plasma	nicotinamide	1.0[0.8: 1.3]	7.3×10^{-1}	21	0.0%	3.8%	12	0.0%	3.0%	1
saliva	X - 12776	1.0[0.8: 1.3]	7.4×10^{-1}	8	0.0%	17.9%	8	0.0%	0.0%	3
urine	creatinine	1.0[0.8: 1.4]	7.4×10^{-1}	15	0.0%	5.7%	3	0.0%	0.0%	0
urine	X - 18241	1.0[0.7: 1.2]	7.4×10^{-1}	19	0.0%	0.0%	9	0.0%	8.3%	0
urine	3-aminoisobutyrate	1.0[0.8: 1.2]	7.5×10^{-1}	11	0.0%	9.1%	7	0.0%	9.5%	0
plasma	X - 14473	1.0[0.8: 1.3]	7.5×10^{-1}	7	0.0%	9.5%	4	0.0%	33.3%	0
plasma	arachidate (20:0)	1.0[0.8: 1.2]	7.6×10^{-1}	15	0.0%	3.8%	7	0.0%	14.3%	0
urine	1,5-anhydroglucitol (1,5-AG)	1.0[0.8: 1.2]	7.6×10^{-1}	18	0.0%	3.9%	3	0.0%	0.0%	0
saliva	phenylalanine	1.0[0.8: 1.3]	7.6×10^{-1}	6	0.0%	33.3%	5	0.0%	40.0%	1
saliva	N-acetylaspartate (NAA)	1.0[0.8: 1.2]	7.9×10^{-1}	10	0.0%	6.7%	8	0.0%	25.0%	0
urine	2-aminophenol sulfate	1.0[0.8: 1.3]	7.9×10^{-1}	8	0.0%	25.0%	6	0.0%	13.3%	0
saliva	X - 19839	1.0[0.8: 1.3]	7.9×10^{-1}	15	0.0%	6.7%	7	0.0%	19.0%	0
urine	urea	1.0[0.8: 1.3]	8.0×10^{-1}	9	0.0%	2.8%	0	0.0%	—	0
urine	O-sulfo-L-tyrosine	1.0[0.8: 1.3]	8.0×10^{-1}	6	0.0%	6.7%	9	0.0%	0.0%	1
plasma	X - 11299	1.0[0.8: 1.3]	8.2×10^{-1}	13	0.0%	0.0%	12	0.0%	3.0%	0
urine	X - 19808	1.0[0.8: 1.2]	8.2×10^{-1}	5	0.0%	10.0%	2	0.0%	0.0%	0
plasma	oleoylcarnitine	1.0[0.8: 1.3]	8.2×10^{-1}	5	0.0%	40.0%	2	0.0%	100.0%	0
urine	phenol sulfate	1.0[0.8: 1.3]	8.3×10^{-1}	5	0.0%	30.0%	6	0.0%	40.0%	0
plasma	X - 19807	1.0[0.8: 1.2]	8.3×10^{-1}	6	0.0%	46.7%	4	0.0%	66.7%	0
plasma	gamma-glutamyltyrosine	1.0[0.7: 1.3]	8.3×10^{-1}	10	0.0%	20.0%	7	0.0%	19.0%	0
plasma	isoleucine	1.0[0.8: 1.4]	8.3×10^{-1}	8	0.0%	10.7%	6	0.0%	26.7%	0
plasma	arginine	1.0[0.8: 1.2]	8.3×10^{-1}	19	0.0%	3.5%	6	0.0%	6.7%	0
plasma	cholate	1.0[0.8: 1.2]	8.4×10^{-1}	10	0.0%	13.3%	8	0.0%	3.6%	0
urine	X - 12211	1.0[0.8: 1.3]	8.5×10^{-1}	10	0.0%	31.1%	2	0.0%	0.0%	0
urine	3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	1.0[0.8: 1.3]	8.5×10^{-1}	15	0.0%	4.8%	13	0.0%	5.1%	1
saliva	X - 14196	1.0[0.8: 1.2]	8.6×10^{-1}	12	0.0%	9.1%	10	0.0%	22.2%	2
urine	X - 12830	1.0[0.8: 1.3]	8.7×10^{-1}	11	0.0%	1.8%	10	0.0%	4.4%	0
plasma	glycocholate sulfate	1.0[0.8: 1.3]	8.8×10^{-1}	11	0.0%	10.9%	12	0.0%	9.1%	0
urine	cyclo(gly-pro)	1.0[0.8: 1.3]	8.8×10^{-1}	13	0.0%	3.8%	5	0.0%	10.0%	1
urine	X - 15636	1.0[0.8: 1.3]	8.8×10^{-1}	10	0.0%	2.2%	6	0.0%	0.0%	1
urine	3-ureidopropionate	1.0[0.8: 1.3]	9.0×10^{-1}	10	0.0%	4.4%	8	0.0%	3.6%	0
urine	3-methylhistidine	1.0[0.8: 1.3]	9.1×10^{-1}	4	0.0%	83.3%	4	0.0%	50.0%	0
plasma	adenosine 5'-monophosphate (AMP)	1.0[0.8: 1.3]	9.3×10^{-1}	20	0.0%	6.8%	17	0.0%	3.7%	0
urine	pseudouridine	1.0[0.8: 1.3]	9.6×10^{-1}	11	0.0%	3.6%	5	0.0%	0.0%	0
urine	X - 12104	1.0[0.8: 1.3]	9.6×10^{-1}	12	0.0%	9.1%	1	0.0%	—	0
saliva	paraxanthine	1.0[0.8: 1.3]	9.8×10^{-1}	8	0.0%	10.7%	7	0.0%	42.9%	0
urine	X - 18486	1.0[0.8: 1.3]	9.9×10^{-1}	11	0.0%	1.8%	5	0.0%	0.0%	0

Figure E.1 Comparison of renal function associations with previous studies. Here I compare the observed univariate associations of metabolites with renal disease with the results of the study by Sekula et al. (2016). 277 blood metabolites overlapped between the studies. For each of them the p-values are shown here.

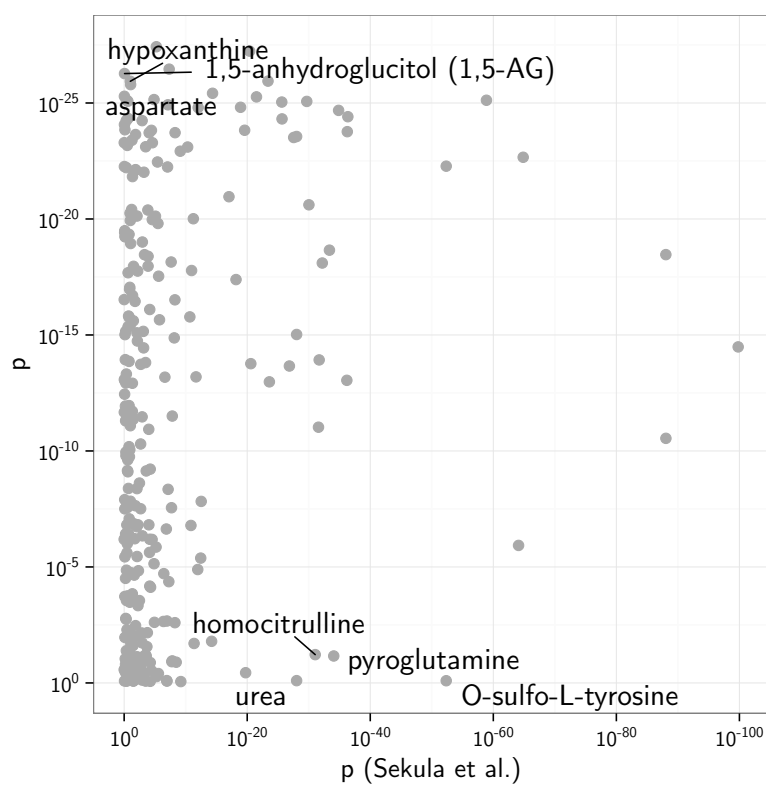


Figure E.2 Principal components of metabolomics measurements. I conducted principal component analysis (PCA) for each of the three fluids after correction for confounding factors and imputation of missing values. Cases are coloured in red, controls in yellow. In each of the fluids, the disease status explains a major portion of variance in the metabolomics profiles.

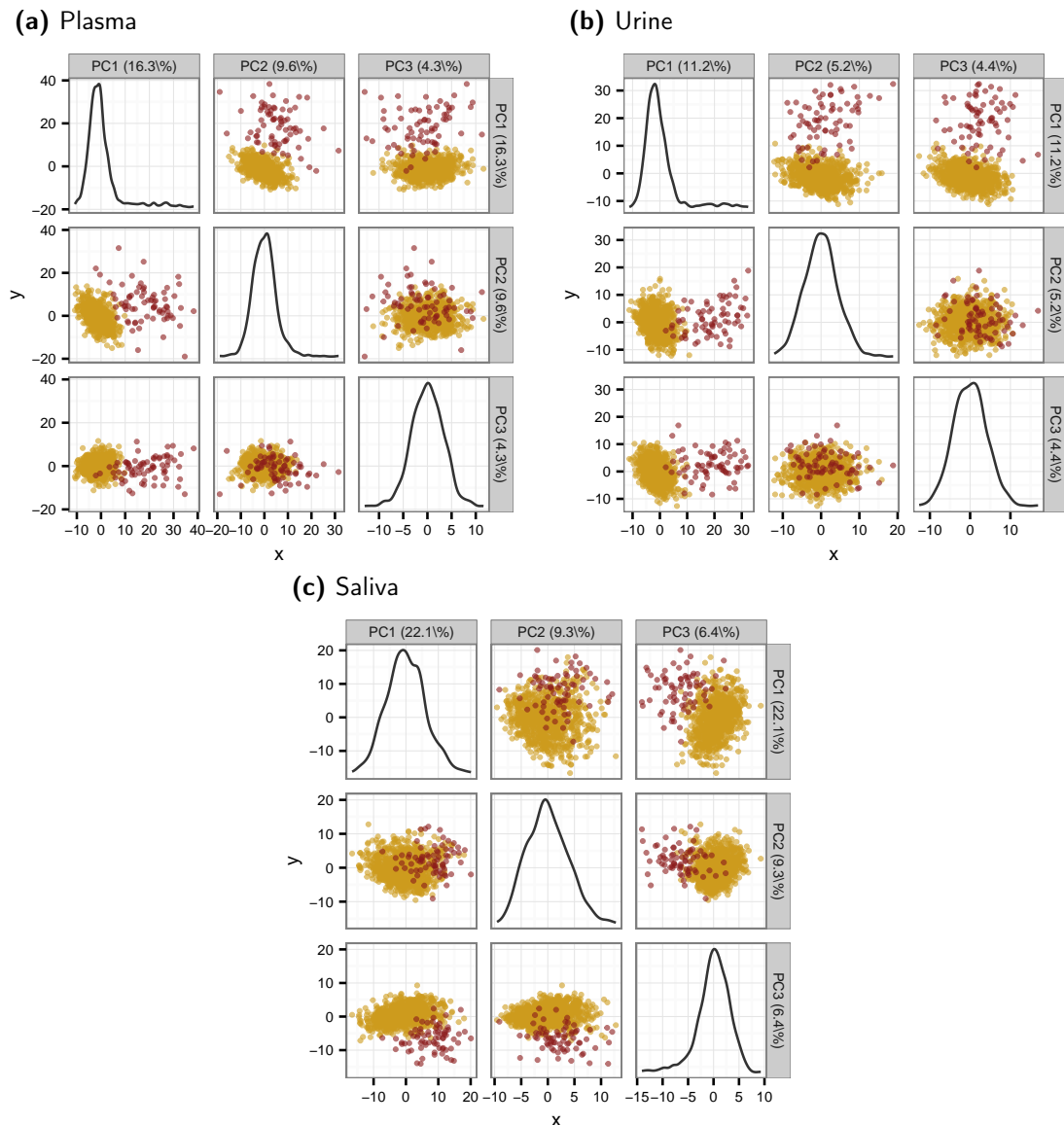


Figure E.3 Metabolites associated with CKD across fluids. Seven metabolites were significantly different between renal disease patients and controls in all three fluids. While associations were consistent across fluids for four metabolites (first row), directions differed between fluids for the remaining three (second row).

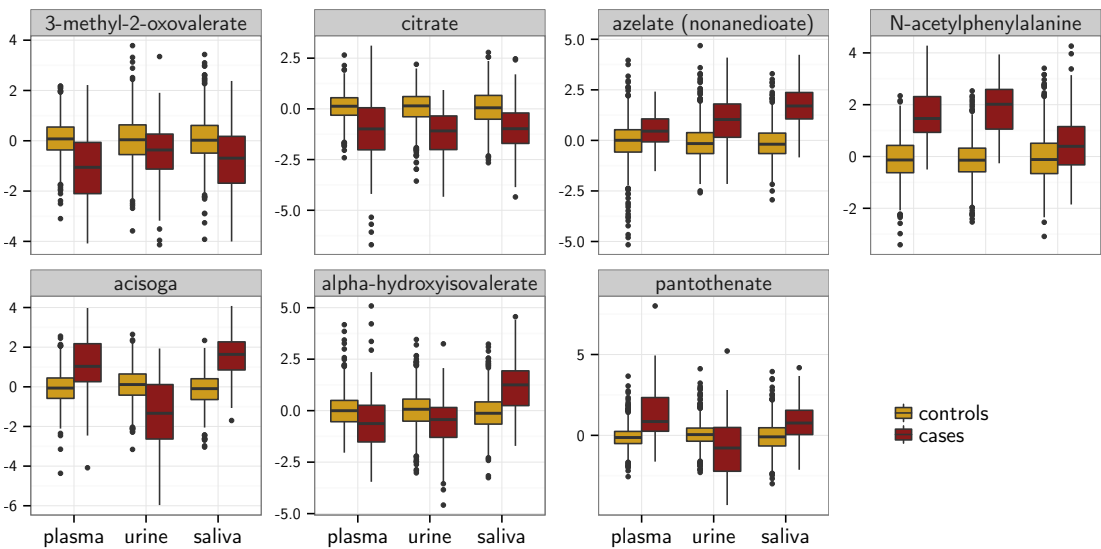
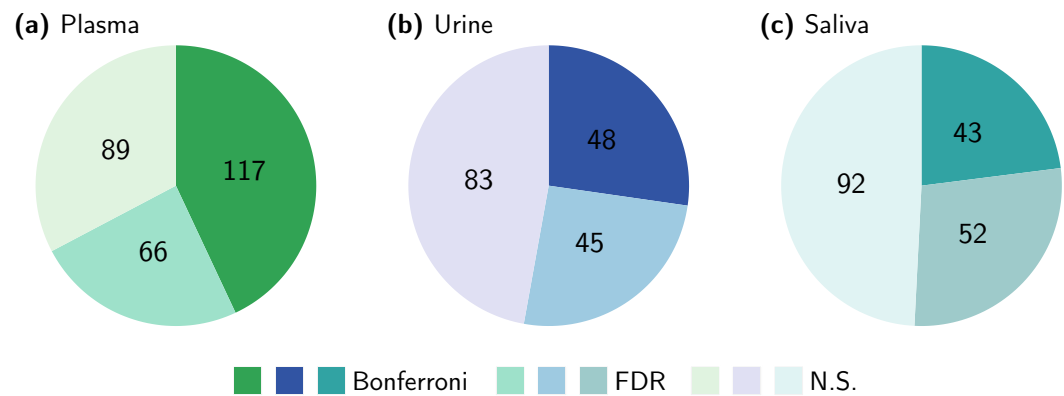


Figure E.4 Rare metabolites differing between CKD cases and controls. I analysed 882 metabolites present in more than 80% of the samples in my main analysis, as network inference relies on full data matrices. However, 782 additional metabolites were detected in the study population, but excluded due to high proportion of missing values. 635 of those were observed in at least 10 samples. Logistic regression models identify 208 of those significantly different between cases and controls.



APPENDIX F

Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model

Table F.1 Omics variables included in the Age-MGM. I included 53 variables from four different 'omics' datasets in the Age-MGM. Degree and clustering coefficients (Clust.) were calculated from the entire model, while the betweenness centrality (Betw.) was calculated for the giant component of the Age-MGM only.

Label	Description	Degree	Clust.	Betw.
cg17861230	PDE4C (chr19, 18343901)	5	40.0%	1.6%
cg02228185	ASPA (chr17, 3379567)	2	0.0%	0.7%
cg25809905	ITGA2B (chr17, 42467728)	0	—	—
GP6	The percentage of FA2B glycan in total IgG glycans	7	38.1%	1.8%
GP14	The percentage of FA2G2 glycan in total IgG glycans	7	47.6%	1.8%
GP15	The percentage of FA2BG2 glycan in total IgG glycans	3	100.0%	0.0%
C-glycosyltryptophan	Amino Acid	9	25.0%	3.9%
citrate	Energy	7	47.6%	0.4%
DHEA-S	Lipid	7	28.6%	5.0%
EPA	Lipid	6	40.0%	5.8%
CMPF	Lipid	6	33.3%	4.2%
10-heptadecenoate	Lipid	6	66.7%	0.2%
urate	Nucleotide	5	40.0%	4.2%
dihomo-linoleate	Lipid	5	80.0%	0.0%
erythritol	Xenobiotics	5	40.0%	1.0%
glutamate	Amino Acid	5	10.0%	8.3%
myristoleate	Lipid	5	80.0%	1.0%
octanoylcarnitine	Lipid	5	80.0%	0.0%
threitol	Carbohydrate	4	33.3%	3.7%
creatinine	Amino Acid	3	66.7%	0.3%
phosphate	Energy	3	33.3%	0.1%
androstendiol-2S	Lipid	3	33.3%	1.2%
androstendiol-S	Lipid	3	66.7%	1.4%
citrulline	Amino Acid	2	0.0%	0.0%
serine	Amino Acid	2	0.0%	0.2%
palmitoyl sphingomyelin	Lipid	2	0.0%	0.2%
aspartate	Amino Acid	1	—	0.0%
creatine	Amino Acid	0	—	—
1,7-dimethylurate	Xenobiotics	0	—	—
ILMN 2062620	NTM2 (chr10, 15187951:15188000)	13	17.9%	6.1%
ILMN 1735124	OXT (chr20, 3001023:3001072)	9	33.3%	6.0%
ILMN 2112638	SVEP1 (chr9, 112168560:112168609)	9	27.8%	10.5%
ILMN 1750018	SEL1L2 (chr20, 13778194:13778243)	8	28.6%	4.5%
ILMN 1684391	PLOD1 (chr1, 11958106:11958155)	7	38.1%	0.3%
ILMN 1708107	DPT (chr1, 166931502:166931551)	7	33.3%	2.6%
ILMN 2065690	GRAMD3 (chr5, 125857223:125857272)	6	53.3%	0.2%
ILMN 1678403	TMEM178 (chr2, 39797960:39798009)	5	60.0%	0.5%
ILMN 1790350	TPRG1 (chr3, 190523591:190523640)	5	20.0%	3.3%
ILMN 1664679	CADM2 (chr3, 86200275:86200324)	4	33.3%	1.0%
ILMN 1727309	FAM82A2 (chr15, 38815707:38815756)	4	16.7%	0.2%
ILMN 1732410	SLC16A9 (chr10, 61080824:61080873)	4	33.3%	2.4%
ILMN 1782069	TRAK1 (chr3, 42228905:42228954)	4	50.0%	0.2%
ILMN 1803686	ADA (chr20, 42681856:42681905)	4	33.3%	2.9%
ILMN 2344283	FMO3 (chr1, 169353283:169353332)	4	50.0%	0.3%
ILMN 1749962	NCAM2 (chr21, 21832517:21832566)	3	100.0%	0.0%
ILMN 1689431	APCDD1L (chr20, 56467867:56467916)	2	100.0%	0.0%
ILMN 1749540	RBM20 (chr10, 112588928:112588977)	2	0.0%	1.5%
ILMN 1814397	EPB42 (chr15, 41276902:41276951)	2	0.0%	2.1%
ILMN 2297864	MTMR14 (chr3, 9714444:9714493)	2	100.0%	0.0%
ILMN 1793543	C1orf51 (chr1, 148525842:148525891)	1	—	0.0%
ILMN 1657087	ZNF533 (chr2, 180015130:180015179)	0	—	—
ILMN 1727833	KIF19 (chr17, 69858798:69858847)	0	—	—
ILMN 1794552	GAP43 (chr3, 116922460:116922509)	0	—	—

Table F.2 Phenotypes included in the Age-MGM. Additional to the ‘omics’ markers, 92 phenotypes were included in the Age-MGM. The table indicates the population average and standard deviation for all continuous variables and the prevalence for all dichotomous phenotypes. Degree, clustering coefficients (Clust.) were calculated from the entire model, where betweenness centrality (Betw.) was calculated from the giant component Age-MGM only

Node	Average	Degree	Clust.	Betw.
ALAT (alanine amino transferase)	28.6 ± 12.6	1	—	0.0%
FEV1 (forced expiratory volume in 1 second))	2.5 ± 0.6	8	0.4%	0.0%
FVC (forced vital capacity)	3.2 ± 0.6	7	0.5%	0.0%
GGT (gamma-glutamyltransferase)	25.7 ± 21.6	2	0.0%	0.0%
activity (home)	330 (71.1%)	1	—	—
activity (leisure)	293 (63.1%)	1	—	—
activity (work)	296 (63.8%)	1	—	0.0%
age	59.0 ± 9.4	27	0.1%	0.5%
albumin	44.2 ± 3.6	2	0.0%	0.0%
android fat mass	2284.4 ± 1022.1	21	0.4%	0.1%
android gynoid ratio	1.0 ± 0.2	15	0.4%	0.0%
android lean mass	3076.1 ± 550.4	13	0.7%	0.0%
ankle swelling	24 (5.3%)	0	—	—
antidepressants	71 (17.3%)	1	—	—
anxiety stress disorder	68 (14.7%)	0	—	—
arthritis	183 (39.3%)	4	0.3%	0.0%
arthritis (other)	57 (13.0%)	2	1.0%	0.0%
asthma	69 (13.5%)	1	—	0.0%
bilirubin	9.0 ± 4.4	0	—	—
birthweight	2381.6 ± 545.0	1	—	0.0%
BMI	26.6 ± 4.9	18	0.4%	0.1%
low back pain	47 (11.5%)	0	—	—
cancer	62 (12.2%)	0	—	—
chest pain	99 (19.4%)	0	—	—
HDL cholesterol	1.8 ± 0.5	5	0.4%	0.0%
high cholesterol	138 (27.1%)	0	—	—
total cholesterol	5.6 ± 1.1	4	0.3%	0.0%
chronic bronchitis	31 (7.0%)	0	—	—
chronic widespread pain	102 (24.9%)	2	0.0%	0.0%
DBP (diastolic blood pressure)	79.3 ± 9.8	2	1.0%	0.0%
depression	50 (10.8%)	1	—	—
dizziness	24 (4.7%)	0	—	—
eGFR (estimated glomerular filtration rate)	81.8 ± 15.0	7	0.3%	0.0%
father overweight	49 (10.8%)	2	1.0%	0.0%
glucose	4.9 ± 0.5	0	—	—
grip strength	26.9 ± 6.2	6	0.5%	0.0%
gynoid fat mass	4712.2 ± 1392.6	12	0.8%	0.0%
gynoid lean mass	6195.9 ± 934.7	14	0.6%	0.0%
hearing	177 (38.0%)	0	—	—
heart murmur	25 (5.4%)	0	—	—
height	161.4 ± 5.8	12	0.4%	0.0%
hip	101.5 ± 9.8	12	0.7%	0.0%
hip BMC (hip bone mineral content)	32.3 ± 8.6	10	0.6%	0.0%
hip BMD (hip bone mineral density)	0.9 ± 0.2	7	0.4%	0.1%
hypertension	120 (23.5%)	2	1.0%	0.0%
irregular heartbeats	25 (5.4%)	0	—	—
learn gadgets	142 (28.6%)	0	—	—
loss of vision/speech	31 (6.7%)	0	—	—
memory (clothes)	30 (6.0%)	0	—	—
memory (keys)	191 (38.5%)	0	—	—
memory (loss)	168 (36.3%)	2	1.0%	—
memory (medication)	54 (10.9%)	0	—	—
memory (month)	59 (11.9%)	0	—	—
memory (rate)	325 (65.7%)	2	1.0%	—
memory (worry)	170 (34.3%)	3	0.3%	—
memory (worsen)	135 (27.2%)	1	—	—
migraine	79 (17.1%)	0	—	—
mother overweight	79 (17.4%)	0	—	—
nutrition (total energy)	1893.7 ± 575.1	0	—	—
nutrition (fruit and vegetable)	0.7 ± 2.1	3	1.0%	0.0%
nutrition (high alcohol)	0.1 ± 1.4	1	—	0.0%
nutrition (traditional english)	0.3 ± 1.5	1	—	0.0%
nutrition (dieting)	0.3 ± 1.3	1	—	0.0%
nutrition (low meat)	-0.2 ± 1.3	2	0.0%	0.0%
osteoarthritis	115 (24.0%)	4	0.3%	0.1%
pain (any)	89 (19.6%)	7	0.8%	—
pain (back)	51 (11.2%)	7	0.8%	—
pain (hand)	47 (10.4%)	6	0.9%	—
pain (hip)	43 (9.5%)	6	0.9%	—
pain (knee)	54 (11.9%)	7	0.8%	—
pain (neck)	32 (7.0%)	6	0.9%	—
pain (other)	22 (4.8%)	5	1.0%	—
painkillers	53 (12.9%)	1	—	0.0%
palpitations	37 (8.1%)	0	—	—
rheumatoid arthritis	25 (4.9%)	1	—	0.0%
SBP (systolic blood pressure)	131.2 ± 17.7	4	0.3%	0.0%
smoking	52 (10.2%)	0	—	—
total spine BMC (spine bone mineral content)	59.0 ± 12.7	7	0.6%	0.0%
total spine BMD (spine bone mineral density)	1.0 ± 0.2	4	0.8%	0.0%
stiffness	24 (5.3%)	4	1.0%	—
thyroid underactive	40 (7.8%)	0	—	—
telomere length	3.6 ± 0.6	0	—	—
total fat mass	6997.3 ± 2298.1	13	0.8%	0.0%
total lean mass	9272.5 ± 1436.6	14	0.6%	0.0%

continued on next page ...

total protein	70.0 ± 4.4	1	—	0.0%
triglycerides	0.0 ± 0.2	8	0.4%	0.0%
varicose veins	98 (19.2%)	0	—	—
visceral fat mass	1242.4 ± 566.8	18	0.4%	0.0%
visceral lean mass	1785.7 ± 288.3	12	0.8%	0.0%
waist	80.1 ± 10.4	11	0.7%	0.0%
waist/hip	0.8 ± 0.1	7	0.9%	0.0%
weight	69.4 ± 13.7	16	0.6%	0.0%

Table F.3 Principal component-derived dietary patterns. Food frequency questionnaire (FFQ) items were grouped in 53 food grouped and common patterns were extracted using PCA. The first five components with the proportion of variance explained are listed here.

Diet pattern	Variance explained	High intakes ¹	Low intakes ²
Fruit and vegetable	8.2 %	Fruit, allium and cruciferous vegetables	Fried potatoes
High alcohol	3.9 %	Beer, wine and allium vegetables	High fibre breakfast cereals and fruit
Traditional English	3.6 %	Fried fish and potatoes, meats, savoury pies and cruciferous vegetables	
Dieting	3.3 %	Low-fat dairy products, low-sugar soda	Butter and sweet baked products
Low meat	3.2 %	Baked beans, pizza and soy foods	Meat, other fish and seafood, and poultry

¹ Food frequency questionnaire items with factor loadings ≥ 0.20

² Food frequency questionnaire items with factor loadings ≤ 0.20

Figure F.1 Pairwise correlations in the Age-MGM. Pairwise associations between variables were assessed using linear and logistic regression models, respectively. Edges in this graph represent associations with false discovery rate (FDR) < 5 %, edge colours represent the sign of the beta (blue for negative betas, red for positive). Nodes were positioned and coloured accordingly to the Age-MGM (Figure 9.3).

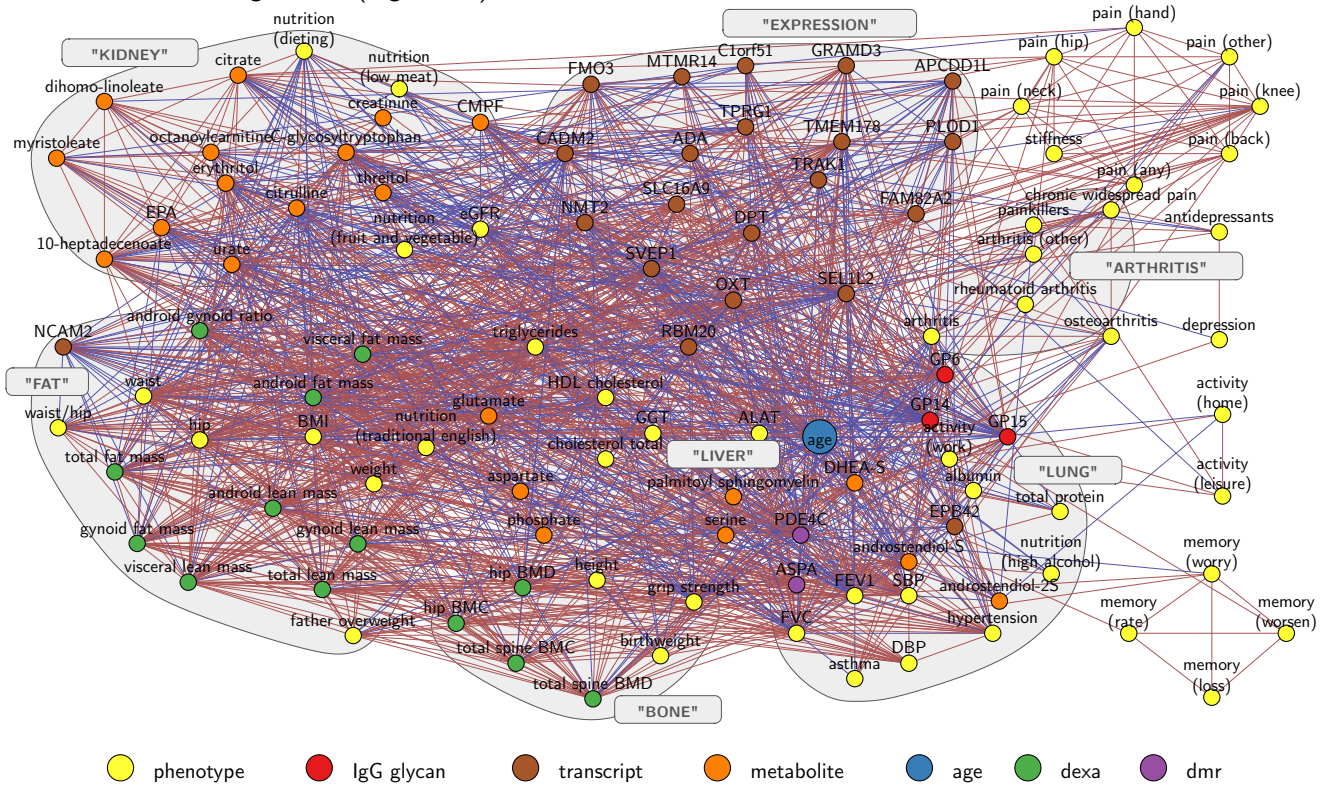


Figure F.2 Connections within and between groups of variables of the Age-MGM. Variables from the Age-MGM (Figure 9.3) were grouped according to the dataset they originate from. Node labels indicate the number of variables in each group, while edge labels indicate the total number of edges between groups.

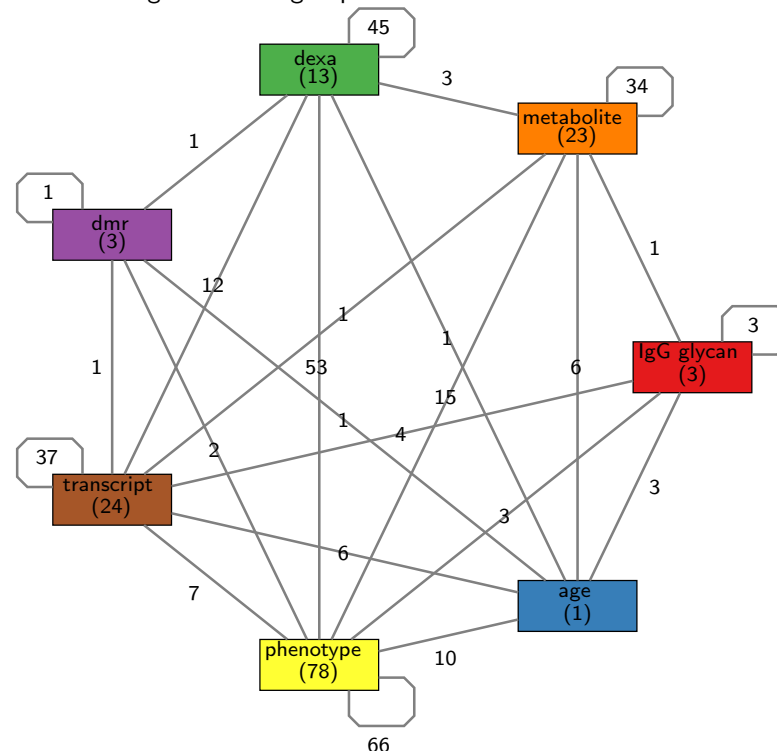
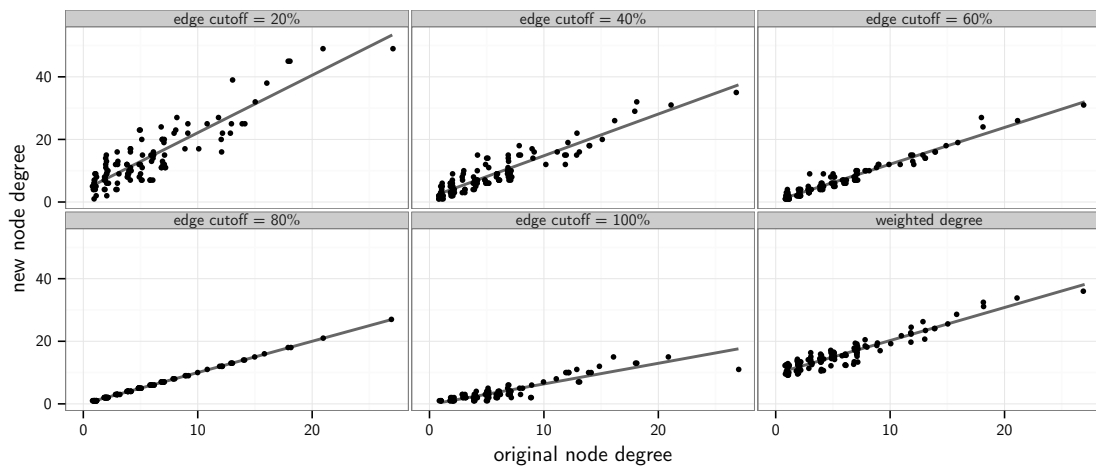
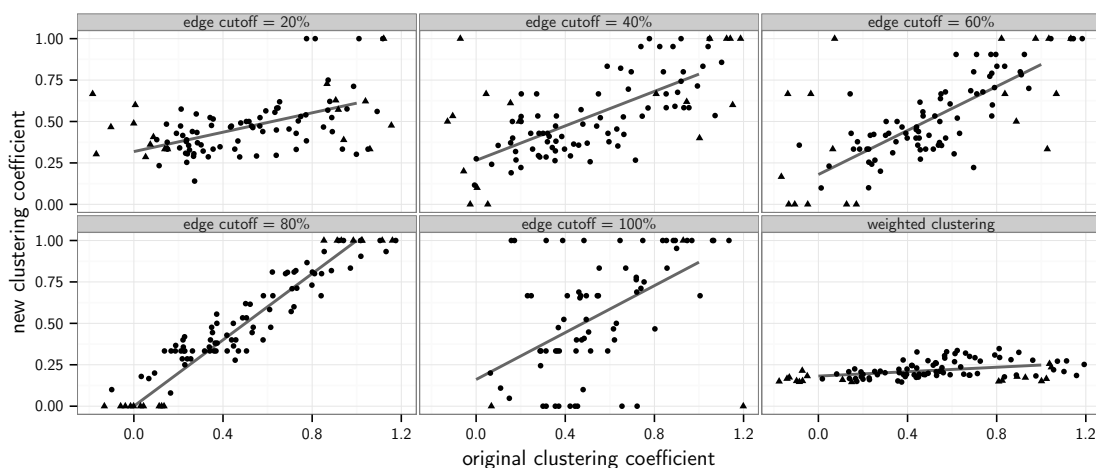


Figure F.3 Stability of Age-MGM with respect to edge cut-off. To assess the dependence of my results on the cut-off for edge inclusion, I compared measures of node centrality and cluster assignments for networks inferred with varying edge inclusion cut-offs (ranging from 20 % to 100 %) as well as a weighted network. (a) Node degrees and (b) clustering coefficients correlate well between the different networks despite their very different sizes (e.g. 969 edges in network with cut-off 20 % compared to 185 edges when using cut-off 100 %). (c) The adjusted RAND index was used to assess similarity of module assignments between networks (where a RAND index of 1.0 indicates identical module assignments and low values indicate disparate modules). It is above 0.4 for all edge inclusion cut-offs, much higher than 1000 randomly sampled module assignments (empirical p-value $<1.0 \times 10^{-3}$), thus illustrating the stability of the overall network structure against various choices for edge cut-offs.

(a) Node degree



(b) Clustering coefficients



(c) Module assignments

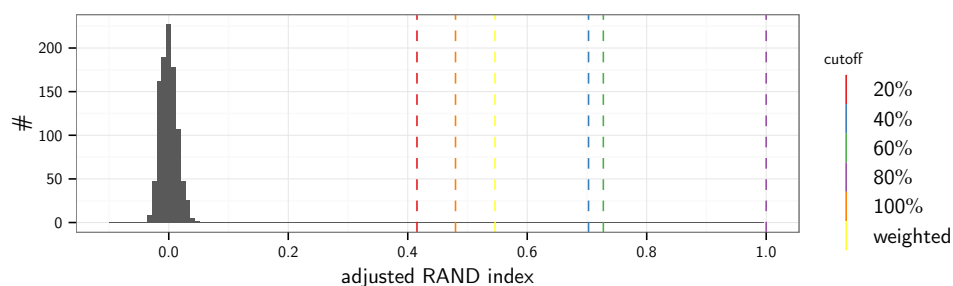
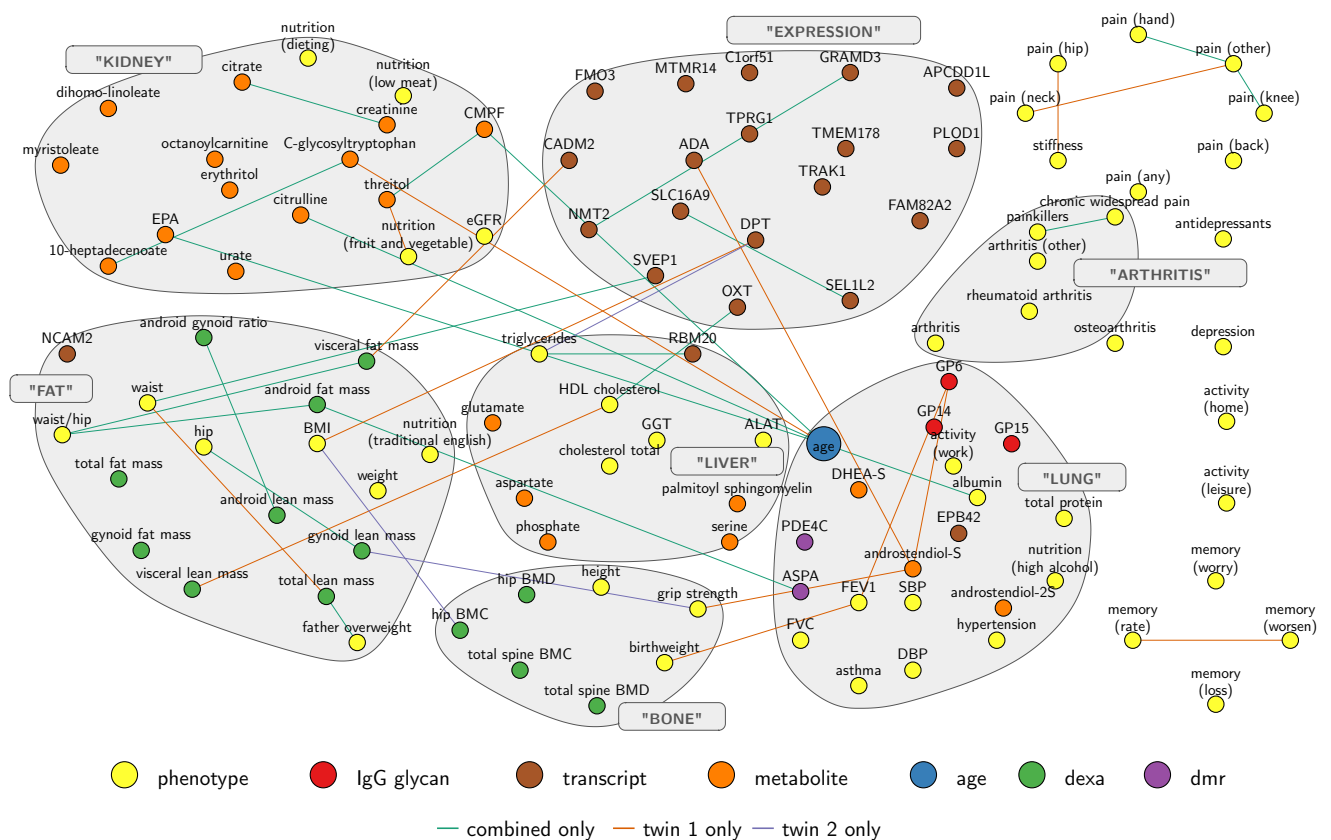


Figure F.4 Stability of the Age-MGM across datasets. stability of the Age-MGM, I additionally inferred two separate models, *twin 1* and *twin 2*, from two disjoint subsets of the data, each of them containing one member of each twin pair. (a) Illustration of the Age-MGM showing only edges that are exclusive to the model based on the combined dataset (green), the *twin 1* (orange) or the *twin 2* (purple) models, respectively. The network is very stable between the datasets, with only 21 edges being unique to the original model. (b) The 21 edges that are unique to the full Age-MGM are on average selected in 78 % of the subsamples of the *twin 1* and *twin 2* models, which demonstrates that they were excluded because they just missed the edge-inclusion cut-off of 80 %, possibly due to a lack of power

(a) Age-MGM with edges exclusive to twin 1 and twin 2 models



(b) Weights of edges unique to the complete Age-MGM

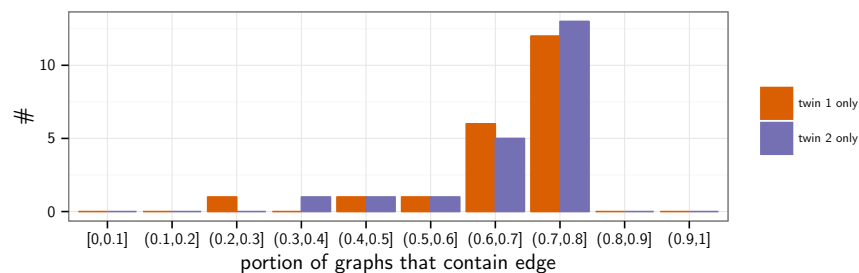


Table F.4 Stability of the Age-MGM against variable selection procedure. To assess the influence variable selection prior to network inference on my results I inferred a second model from the same dataset but without pre-selecting metabolomics variables, thus more than doubling the number of nodes. I then restricted this much larger network to the nodes of the Age-MGM. While the underlying larger model is very different from the original one in terms of size, sparsity and included variables, the network modules were stable across both networks (adjusted RAND index 0.57). Detailed module memberships are listed for this network and the original age- mgm. Variables in *italics* have degree 0 in the large network and consequently not part of any cluster.

Module	Members in Age-MGM	Members in Large-MGM
EXPRESSION	CADM2, TMEM178, PLOD1, APCDD1L, DPT, FAM82A2, SLC16A9, OXT, SEL1L2, TRAK1, TPRG1, C1orf51, ADA, NMT2, GRAMD3, SVEP1, MTMR14, FMO3	CADM2, TMEM178, PLOD1, APCDD1L, FAM82A2, TRAK1, ADA, NMT2, GRAMD3, SVEP1, MTMR14, FMO3, SLC16A9, TPRG1, C1orf51
LUNG	FEV1, FVC, GP14, GP15, GP6, EPB42, androstendiol-S, DHEA-S, androstendiol-2S, activity (work), age, <i>albumin, asthma, ASPA, PDE4C, DBP, hypertension, nutrition (high alcohol), SBP, total protein</i>	FEV1, FVC, GP14, GP15, GP6, OXT, RBM20, SEL1L2, EPB42, androstendiol-S, DHEA-S, activity (work), age, ASPA, PDE4C, DBP, hypertension, SBP
ARTHRITIS	arthritis, arthritis (other), chronic widespread pain, osteoarthritis, <i>painkillers</i> , rheumatoid arthritis	arthritis, arthritis (other), chronic widespread pain, osteoarthritis, rheumatoid arthritis
BONE	birthweight, grip strength, height, hip BMC, hip BMD, total spine BMC, total spine BMD	DPT, NCAM2, android fat mass, android gynoid ratio, android lean mass, BMI, gynoid fat mass, hip, total fat mass, total lean mass, visceral fat mass, visceral lean mass, waist, waist/hip, weight
FAT	NCAM2, android fat mass, android gynoid ratio, android lean mass, BMI, <i>father overweight</i> , gynoid fat mass, gynoid lean mass, hip, <i>nutrition (traditional English)</i> , total fat mass, total lean mass, visceral fat mass, visceral lean mass, waist, waist/hip, weight	DPT, NCAM2, android fat mass, android gynoid ratio, android lean mass, BMI, gynoid fat mass, hip, total fat mass, total lean mass, visceral fat mass, visceral lean mass, waist, waist/hip, weight
LIVER	ALAT, GGT, RBM20, phosphate, aspartate, serine, glutamate, palmitoyl sphingomyelin, HDL cholesterol, cholesterol total, triglycerides	ALAT, GGT, phosphate, aspartate, serine, glutamate, androstendiol-2S, palmitoyl sphingomyelin, HDL cholesterol, cholesterol total, triglycerides
KIDNEY	creatinine, <i>citrate, urate, citrulline, dihomolnoleate, EPA, erythritol, CMPF, myristoleate</i> , C-glycosyltryptophan, <i>octanoylcarnitine, 10-heptadecenoate</i> , threitol, eGFR, <i>nutrition (fruit and vegetable), nutrition (dieting), nutrition (low meat)</i>	creatinine, urate, erythritol, C-glycosyltryptophan, threitol, eGFR

APPENDIX G

Published Articles

G.1 Integration of 'omics' data in aging research: from biomarkers to systems biology

REVIEW

Integration of 'omics' data in aging research: from biomarkers to systems biology

Jonas Zierer,^{1,2} Cristina Menni,¹ Gabi Kastenmüller^{1,2} and Tim D. Spector¹

¹Department of Twins Research and Genetic Epidemiology, Kings College London, London, United Kingdom

²Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany

Summary

Age is the strongest risk factor for many diseases including neurodegenerative disorders, coronary heart disease, type 2 diabetes and cancer. Due to increasing life expectancy and low birth rates, the incidence of age-related diseases is increasing in industrialized countries. Therefore, understanding the relationship between diseases and aging and facilitating healthy aging are major goals in medical research. In the last decades, the dimension of biological data has drastically increased with high-throughput technologies now measuring thousands of (epi) genetic, expression and metabolic variables. The most common and so far successful approach to the analysis of these data is the so-called reductionist approach. It consists of separately testing each variable for association with the phenotype of interest such as age or age-related disease. However, a large portion of the observed phenotypic variance remains unexplained and a comprehensive understanding of most complex phenotypes is lacking. Systems biology aims to integrate data from different experiments to gain an understanding of the system as a whole rather than focusing on individual factors. It thus allows deeper insights into the mechanisms of complex traits, which are caused by the joint influence of several, interacting changes in the biological system. In this review, we look at the current progress of applying omics technologies to identify biomarkers of aging. We then survey existing systems biology approaches that allow for an integration of different types of data and highlight the need for further developments in this area to improve epidemiologic investigations.

Key words: data integration; graphical models; high-throughput data; omics; systems biology.

Introduction

Aging is often described as the progressive accumulation of changes with time leading to a loss of physiological aptitude and fertility, an

increased susceptibility to disease and ultimately to death (Harman, 1988, 2001; Kirkwood & Austad, 2000; Vijg & Suh, 2005; López-Otín *et al.*, 2013). Despite considerable effort and the development of many theories, the underlying process is still largely unknown (Kirkwood & Austad, 2000; Weinert & Timiras, 2003; Rattan, 2006).

Researchers distinguish between chronological and biological age. Chronological age is defined as the absolute time that an individual lives. In contrast, biological age is a broader concept that takes the individual physical and mental health into account, thus capturing individual differences of the aging process. Most aging studies search for associations of chronological age with clinical and molecular phenotypes (Warming *et al.*, 2002). However, several studies used phenotypes, such as lung function, grip strength or bone mineral density, as proxies to investigate molecular changes in biological aging (Jackson *et al.*, 2003; Bell *et al.*, 2012; Levine, 2013). Researchers also investigated reasons of retarded biological aging and longevity by comparing centenarians with younger controls (Biagi *et al.*, 2012; Sebastiani *et al.*, 2012).

The life expectancy in the UK increased by 5.3 years for men and 4.7 for women over the last two decades and is predicted to further increase in the next twenty years (Oeppen & Vaupel, 2002; Office for National Statistics 2014). With increasing life expectancy, age-related diseases are expected to rise dramatically (700 000 people suffered from dementia in 2000, 800 000 in 2012 and approximately 1 million people will be affected by dementia in 2021 (Alzheimer's Society 2014)) with major impacts on healthcare costs. Thus, a better understanding of aging and its influence on disease is a long term public health goal and a hot topic of current medical research.

Omics technologies provide valuable tools to study aging on the molecular level. Reductionist data analyses, testing the measured variables separately for association with age, have been extensively applied. Such studies successfully identified hundreds of epigenetic mutations, gene expression levels, metabolite concentrations to be linked with chronological and/or biological age (see below for details). Even though these results improved our understanding of aging as a complex phenotype, the mechanisms underlying these associations and the impact of interactions between different biological entities remain elusive in most cases. In contrast to reductionist approaches, systems biology aims to analyse all components of a biological process simultaneously taking into account their interactions and their intrinsic hierarchical structure (Ideker *et al.*, 2001; Barabási & Oltvai, 2004). With more and more high-throughput data becoming available, systems biology has led to many new methods and their successful application on age and age-related phenotypes (as outlined below).

In this review, we will briefly summarize the current progress in 'omics' technologies and their application in aging research. We will then highlight some problems of the reductionist approach and discuss how these may be overcome using systems biology. We present a selection of statistical methods used in systems biology along with their current and possible future applications in the field of aging research to move from biomarkers of aging to a more holistic understanding of the aging process.

Correspondence

Professor Tim D. Spector, Department of Twins Research and Genetic Epidemiology, King's College London, St Thomas' Campus, Lambeth Palace Road, London SE1 7EH, United Kingdom. Tel.: +44 (0) 207 188 6765; fax: +44 (0) 207 188 6761; e-mail: tim.spector@kcl.ac.uk

Accepted for publication 20 June 2015



Omics and aging

New technologies allow the measurement of 'omics' data and numerous association studies have been conducted. Valdes *et al.* (2013) thoroughly reviewed the application of these technologies to identify molecular markers of aging from each omics level. Therefore, the following section will only briefly highlight some key results and concentrate on recent findings.

Genomics

Genomics was the first omics field for which high-throughput measurements became available. Current chips are able to measure up to 5 million single nucleotide polymorphisms (SNPs) (Ha *et al.*, 2014). Today, next-generation sequencing technology is slowly replacing the chip technology as the cost of sequencing has dropped below \$0.10 per million bp (Liu *et al.*, 2012). Thus, gene variation is nowadays often available at single nucleotide resolution.

While aging (or rather longevity) itself was found to be only about 20% heritable (Murabito *et al.*, 2012), many age-related diseases are highly heritable. For instance, Alzheimer's disease (AD) shows a heritability above 70% (Gatz *et al.*, 2006) and osteoarthritis (Ishimori *et al.*, 2010) or cataract show 50% heritability (Hammond *et al.*, 2001).

The GenAge database contains about 300 human candidate genes for aging based on homology with model organisms (Tacutu *et al.*, 2013). Sebastiani *et al.* (2012) recently published a refined model consisting of 281 SNPs to distinguish between centenarians and younger controls in a cohort of 1715 people. One of these SNPs is located in ApoE, which is so far the only gene that has been reliably associated with longevity at genomewide significance level (Deelen *et al.*, 2011; Nebel *et al.*, 2011). Common genetic variants at this locus have been associated with accelerated aging and cognitive decline (Johnson, 2006; Davies *et al.*, 2014), possibly by increasing the risk for coronary artery disease, stroke and AD (Smith, 2002). Even though some studies provided evidence that mutations of FOXO transcription factors are related to longevity (Willcox *et al.*, 2008; Flachsbarth *et al.*, 2009), as well, GWASs failed to replicate this at the level of genomewide significance.

Epigenomics

Epigenomics describes the study of heritable changes in the genome that are not caused by DNA sequence mutations (Lodish, 2013). The most common epigenetic mechanism is DNA methylation, which is known to often silence gene expression. In contrast to the genome, which is the same in all cells, the epigenome is an important factor of cell differentiation leading to profound epigenetic differences across different cell types (Meissner, 2010). The current methylation chip by Illumina measures over 485 000 methylation sites and covers 99% of all RefSeq genes (Illumina 2011). However, it covers less than 10% of variable regions (Ziller *et al.*, 2013).

The epigenome is influenced by environmental and lifestyle factors (Nakajima *et al.*, 2010; Alegría-Torres *et al.*, 2011; Breitling *et al.*, 2011) and is associated with many complex diseases such as neurodegenerative disorders (reviewed by Portela & Esteller, 2010) and cancer (Ehrlich, 2002; Horvath, 2013). Nearly 500 differentially methylated regions were found to be associated with chronological age and age-related phenotypes such as lung function, cholesterol levels and maternal longevity (Bell *et al.*, 2012). A recent study by Weidner *et al.* (2014) showed that methylation patterns of just three sites are sufficient to predict

chronological age. Thus, many of the previously identified methylation sites might not be independently associated with age. Interestingly, variation in methylation with age is consistent across several tissues and cell types (Horvath, 2013). Together, they form a global pattern of hypomethylation in repetitive sequences, hypermethylation in promoter regions and higher intercell variability (Cevenini *et al.*, 2008; Bacalini *et al.*, 2014). Besides DNA methylation, other epigenetic changes, such as histone methylation and acetylation, have been found to be associated with longevity in model organisms (Dang *et al.*, 2009; Greer *et al.*, 2010). Investigating these modifications in humans could shed light on so far unknown mechanisms of aging.

Transcriptomics

Genes are transcribed into RNA molecules, which are further processed in a tightly controlled process. The entirety of the RNA transcripts is referred to as transcriptome. It can be divided in coding RNAs, which are further translated in proteins, and noncoding RNAs, which perform various functions, such as regulation of gene expression (Eddy, 2001). Transcript abundances can be measured by either chips or sequencing methods.

Similar to the epigenome, gene expression was shown to dramatically change with age. A pioneer study comparing postmortem human frontal cortex tissue samples between 30 individuals of different ages yielded 463 differentially expressed genes (Lu *et al.*, 2004). Despite the small sample size, results were replicated in subsequent experiments. Four years later, Berchtold *et al.* (2008) identified several thousand age-related changes in gene expression in four different brain tissues. Later studies by different groups identified profound changes in the transcriptome with age in further tissues, such as skin, adipose tissue ($N = 865$) (Glass *et al.*, 2013) and kidney ($N = 134$) (Rodwell *et al.*, 2004). Most of these changes did not overlap in different tissues. A meta-analysis across different species and tissues revealed only 73 genes consistently associated with age (de Magalhães *et al.*, 2009). This suggests that most observed age-related changes in the transcriptome are either species and tissue specific or false-positive discoveries (reviewed by Valdes *et al.*, 2013). In their meta-analysis, genes related to immune response and lysosome tended to be overexpressed, while genes related to mitochondria and oxidative phosphorylation were underexpressed in elderly (de Magalhães *et al.*, 2009).

Proteomics

Proteins are translated from coding transcripts. Due to alternative splicing and post-translational protein modifications, the number of proteins is estimated to be two orders of magnitudes higher than the number of genes (Ginsburg & Haga, 2006). However, current proteomic techniques based on immunoassays, protein arrays or mass spectrometry can measure only a small fraction of the proteome (up to 1000 proteins in a sample). The most comprehensive description of the human proteome across various tissues to date consists of 18 097 proteins (19 376 isoforms) collected from ten thousand mass spectrometry experiments (Wilhelm *et al.*, 2014).

Due to these technicalities, 'proteomics' studies in aging research so far focused on smaller sets of proteins and small sample sizes. In an early study of protein abundance in the vastus lateralis muscle, Gelfi *et al.* (2006) observed higher abundance of several proteins involved in aerobic metabolism and a lower abundance of proteins involved in anaerobic metabolism in the elderly. Besides this, six transport proteins were consistently underexpressed in older individuals. However, only 12

samples were analysed in this study without replication. A recent study by our group analysed over 1000 proteins in 200 plasma samples using the SOMAscan assay (Menni *et al.*, 2015). Eleven proteins were found to strongly associate with chronological age as well as age-related phenotypes such as lung function and blood pressure. The results were replicated in an independent cohort.

Even though comprehensive proteomics studies are still missing, proteins are likely to be associated with several age-related diseases. For instance, cardiovascular disease (Mehra *et al.*, 2005) and AD (Swardfager *et al.*, 2010) are consistently associated with elevated levels of pro-inflammatory cytokines.

Post-translational modifications – glycomics

Post-translational modifications are important elements of proteins, which can alter their biochemical properties such as protein structure, binding preferences and enzyme activity. There are many different modifications ranging from addition of small molecules (e.g. acetylation or phosphorylation), over addition of larger molecules such as lipids or sugar chains (e.g. palmitoylation, glycosylation), to the addition of whole proteins (e.g. ubiquitination).

The most common modification is glycosylation, which attaches sugar chains to proteins. The attached oligosaccharides – glycans – are supposed to mainly serve as structural elements of proteins or specific binding sites for other glycans or proteins (Varki *et al.*, 2009). However, glycans are highly diverse and many of them are not yet characterized or annotated. Thus, glycans might have many additional functions. For example, glycans in the gut act as food for microbes (Koropatkin *et al.*, 2012), which could be implicated in immune functions that are important in aging. Recent development allows the high-throughput measurement of glycans of either a single protein or all proteins simultaneously (Royle *et al.*, 2008; Pucić *et al.*, 2011).

The application of this technology on epidemiological cohorts revealed that glycan structures are stable for one individual over time (Gornik *et al.*, 2009) but very diverse within a population (Knezević *et al.*, 2009; Pucić *et al.*, 2011). Differences in glycomes were found to be related with various cancers (Fuster & Esko, 2005; Adamczyk *et al.*, 2012). Recently, Kristic *et al.* (2013) showed that IgG glycans are strongly associated with age: a linear combination of three glycans explained 58% of the observed variance of chronological age (Kristic *et al.*, 2013) in a study of four independent populations with 5117 participants in total.

Metabolomics

Metabolomics investigates the low-molecular-weight molecules in a biological system. The measured molecules are often referred to as metabolites as many of them act as educts, products and intermediates of the cellular metabolism. Currently, the Human Metabolome Database (Wishart *et al.*, 2013) contains more than 40 000 distinct metabolites from different tissues. Similar to proteomics, to date, there is no analytical method available to determine and quantify all metabolites in a single experiment. Current platforms, using either chromatography coupled with mass spectrometry or nuclear magnetic resonance, can measure roughly a thousand metabolites in untargeted settings and a smaller number using predefined targeted approaches. The restriction of the targeted approach comes with the advantages of higher sensitivity, absolute instead of relative quantification and straight-forward compound identification (Patti *et al.*, 2012; Tzoulaki *et al.*, 2014).

In 2008, the first metabolome-wide association study on age analysed the plasma metabolome of 269 individuals using an untargeted approach. The authors found 100 of 300 compounds to correlate with chronological age (Lawton *et al.*, 2008). More recently, larger cohorts were employed to study the association of metabolites and age using both targeted and untargeted metabolomics platforms. Yu *et al.* (2012) analysed 131 targeted metabolites in 2162 individuals from the KORA study, while we analysed 280 untargeted metabolites in 6055 twins from the TwinsUK cohort (Menni *et al.*, 2013b). Both studies identified half of the analysed metabolites to be associated with chronological age. Many of the those metabolites were also found to significantly correlate with age-related phenotypes such as lung function, bone mineral density and cholesterol levels (Menni *et al.*, 2013b), AD (N = 93) (Orešič *et al.*, 2011), cancer (reviewed by Teicher *et al.*, 2012) and type 2 diabetes (N = 100) (Suhre *et al.*, 2010; Menni *et al.*, 2013a). One of those metabolites is C-glycosyltryptophan, a potential degradation product of glycosylated proteins.

Microbiomics

The human microbiome describes the complete set of microbial species (and their genomes) hosted by the human body. The largest microbial community resides in the gut, where microbial cells and their genes outnumber human cells (10:1) and genes (100:1) (Peterson *et al.*, 2009; Zhu *et al.*, 2010; The Human Microbiome Project 2014a). More than 10 000 different species with millions of protein-coding genes were identified by the Human Microbiome Project (Turnbaugh *et al.*, 2007; Peterson *et al.*, 2009; Biagi *et al.*, 2012) and >1000 of these microbes have so far been fully sequenced (The Human Microbiome Project 2014b). Although twin studies have found a modest genetic influence on some phyla, most of the variation is environmental (Goodrich *et al.*, 2014).

The composition of the microbe flora varies a lot across individuals (Turnbaugh *et al.*, 2007; Zhu *et al.*, 2010) and even between different parts of the body (Kong, 2011). It has a huge influence on many biological processes such as immune response, metabolism and disease (Zhu *et al.*, 2010; Grice & Segre, 2012). While the microbiome seems to be relatively stable during adulthood, it changes significantly in later life (Guigoz *et al.*, 2008; Biagi *et al.*, 2010; Claesson *et al.*, 2011). Biagi *et al.* (2010) observed drastic changes in the gut microbiome of centenarians compared with young adults as well as elderly, namely a general loss of diversity and increased abundance of bacilli and proteobacteria. The latter were reported to promote inflammation under certain conditions (Round & Mazmanian, 2009). Similar findings were revealed in other elderly populations, which also considered the dietary and residential situation of elderly patients (Claesson *et al.*, 2012).

Phenomics

Simultaneously with omics data, the dimension of clinical and lifestyle traits, particularly clinically used intermediate traits, keeps increasing. Epidemiological studies collected thousands of clinically relevant phenotypes beyond omics data types. These range from anthropometric measures to health and lifestyle questionnaires (Moayyeri *et al.*, 2013). Collecting high-dimensional clinical data is important to unveil pleiotropy of genes and interactions amongst clinical phenotypes such as comorbidities (Houle *et al.*, 2010). Driven by omics technologies, statistical and bioinformatic methods to analyse high-dimensional data are becoming available. These facilitate the investigation of numerous clinical

phenotypes in parallel, thus defining the new field of *phenomics* (Houle *et al.*, 2010).

Phenomics is especially important for aging research. Dozens of clinical phenotypes, such as Parkinson's (Reeve *et al.*, 2014), AD (McAuley *et al.*, 2009), body mass index, blood pressure (Mungreiphy *et al.*, 2011) and bone mineral density (Warming *et al.*, 2002), as well as lifestyle parameters, such as nutrition (Wieser *et al.*, 2011), smoking and physical activity, are strongly related to age (Harman, 1988; Wang *et al.*, 2009). Composite measures such as the Rockwood frailty index (Rockwood & Mitnitski, 2007) combine several of those clinical traits to form a more homogenous phenotype – frailty – from its diverse appearance. Such frailty measures can be considered as measures for biological age (Mitnitski *et al.*, 2013). Many of these (and other) clinical phenotypes correlate or even depend on each other (McAuley *et al.*, 2009; Baylis *et al.*, 2014). Only extensive collection of data and their joint analysis will help to unveil these dependencies and find causal relationships.

From omics to systems biology

Most of the studies summarized above concentrated on the bivariate associations of age (or age-related diseases) with one type of omics data. However, there are strong interdependencies within and between the different omics data (see Fig. 1).

Correlations can be observed practically between all levels of biological organization. Following the central dogma of molecular biology, genomics, transcriptomics and proteomics are correlated 'by definition'. Furthermore, metabolite concentrations are influenced by genetic variants (Shin *et al.*, 2014) and epigenetic factors (Petersen *et al.*, 2014) mediated through changes in gene expression or enzyme activity. Methylation levels do not only influence the gene expression (Jaenisch & Bird, 2003), but are also correlated with gene variants (Bell *et al.*, 2012) and environmental factors (Breitling *et al.*, 2011). Our group has recently demonstrated that even the microbe composition is partly under host genetic influence (Goodrich *et al.*, 2014). Similarly, all levels of omics data are influenced by genetics as well as by environment and aging. Correlations, however, do not only occur between but also within each type of data. For instance, in genomics linkage disequilibrium, the

correlated occurrence of SNPs is a ubiquitous phenomenon. Transcription factors often coregulate the expression of multiple genes (Allocco *et al.*, 2004), and methylation patterns of neighbouring CpG sites were reported to be correlated (Bell *et al.*, 2012). Metabolites are linked by a network of biochemical reactions, causing strong correlations between them (Krusiek *et al.*, 2011). Even phenotypes often cluster. Comorbidities, the over proportional co-occurrence of diseases, were shown to affect many diseases possibly through shared underlying mechanisms (Goh *et al.*, 2007).

These biological correlations can confound the associations and this is a major issue of current research. For instance, 153 metabolites were found by our group to be associated with age, but subsequent analyses showed that only 22 of them are associated with age independently (Menni *et al.*, 2013b). Similarly, 21 of 24 measured IgG glycans were correlated with age, but only 3 of them explain 58% of the variance (Krstic *et al.*, 2013). The same was found for epigenetic data (Weidner *et al.*, 2014). Huge lists of associations with aging are being unveiled using all kinds of data, but the biologically interesting, causal associations are often obscured by this wealth of results. Approaches taking simultaneously information from all omics levels into account are needed to reconstruct the processes involved in aging on a systems level (Valdes *et al.*, 2013).

Even though high-throughput technologies are advancing and more and more data are becoming available, integration of omics remains a challenging problem. Besides the restricted availability of multi-omics data sets for the same samples, technical limitations hamper the integration process. While genomics and transcriptomics are able to measure the entire set of variants, other omics (e.g. proteomics and metabolomics) measure only a small fraction of all entities. Many high-throughput technologies suffer from considerable technical variation and strong batch effects. Stringent quality control and thorough data normalization are crucial when analysing this type of data. Furthermore, the complexity of the organism has to be taken into account. While the genome is more or less stable, all other levels of omics change between cell types and over time. Many samples, such as whole blood, contain a mixture of different cell types with potentially different epigenomes, transcriptomes (Houseman *et al.*, 2012; Jaffe & Irizarry, 2014). Finally, different organs and cells influence each other. The blood metabolome, for instance, is heavily influenced by processes occurring in the liver or in other organs, and multitissue samples are needed to fully understand these. This in turn is not always feasible in an epidemiological setting as collection of tissues often involves invasive procedures. Nevertheless, data integration is an important and active field of research. A first step of data integration is the integration and joint interpretation of separate results. The Digital Ageing Atlas (Craig *et al.*, 2014) summarizes more than 4000 age-related changes across different technologies to facilitate systems-level analyses of aging.

Introduction to systems biology

The aim of systems biology is to understand the system and its functions as a whole rather than as separate components (Cassman, 2005), with the final objective to mathematically model biological systems and simulate their outcomes. As a first step, the complex interactions and dependencies between these components must be formally described to enable systematic analysis and simulation of the biological system of interest. A technique widely used in systems biology is to translate biological interactions into mathematically well-defined networks (graphs). For instance, metabolites interact in chemical reactions, thus forming a network in which *nodes* describe the metabolic compounds

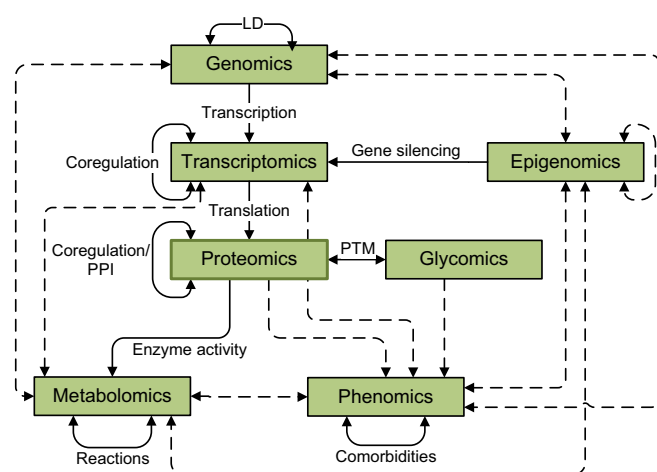


Fig. 1 Interdependencies of omics data: The figure illustrates dependencies which can be observed within almost any omics data set. Solid lines indicate biological processes which cause dependencies, while dashed lines represent observed associations.

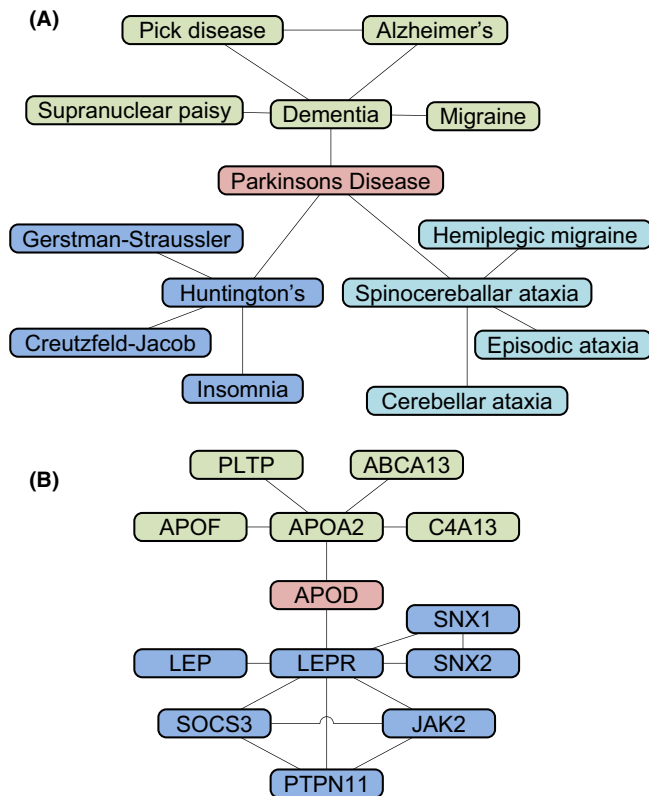


Fig. 2 Topological Properties of Biological Networks (A) is an excerpt from the human disease network (Goh *et al.*, 2007). Nodes represent diseases; these are connected if they are associated with the same gene. Parkinson's disease connects three isolated disease clusters (colours), thus having a low clustering coefficient (0%) and high betweenness (72%). (B) is the close neighbourhood of the ApoD protein in a PPI network from STRING DB (Franceschini *et al.*, 2013) using only experimentally confirmed interactions. ApoD connects two clusters and is, despite the low degree (2) and clustering coefficient (0%), a central node (betweenness centrality: 53%). In contrast, LEPR is central within the blue cluster (degree: 7, clustering: 14%).

and edges indicate chemical reactions. Similarly, transcription factors bind DNA to control gene expression, forming the gene regulatory network (GRN) and interacting proteins build a protein–protein interaction network (PPI) (cf. Fig. 2B). These networks interact, making data integration an important aspect of systems biology. One example for a phenotypic network was created by Goh *et al.* (2007) using diseases as nodes and connecting diseases with shared genetic risk factor by edges (cf. Fig. 2A). By doing so, they showed that many disorders share a set of underlying genetic risk variants and that similar diseases are caused by similar genes.

Graphs can be explored using a variety of established algorithms. One common task is the identification of modules, that is subgraphs in which nodes share certain properties. In biological networks, modules correspond to functional units, such as the glycolysis pathway in the metabolic network. The modules are usually interconnected and together form a hierarchical structure in which the distribution of node degrees – the number of edges per node – follows the power law (Barabási & Oltvai, 2004). Hence, most nodes have only few connections and few nodes have many connections. These highly connected nodes are called hubs (Albert *et al.*, 2000; Jeong *et al.*, 2001). Several other measures exist to describe the topology of networks and topological features of nodes. For example, the clustering coefficient measures how

densely the neighbourhood of a node is connected and thus highlights nodes which are central within a cluster (e.g. LEPR in Fig. 2A). Another measure is the betweenness centrality, which measures the proportion of pairwise shortest paths containing a node. It thus quantifies the importance of a node for connecting other nodes from different modules (e.g. Parkinson's disease in Fig. 2A and APOD in Fig. 2B). The highly connected, central nodes are thought to be key players in the system, connecting several modules and controlling network fluxes. They were shown to be of particular importance for many diseases and survival of the organism (Barabási & Oltvai, 2004; Joy *et al.*, 2005; Yu *et al.*, 2007).

Many software packages for graph analysis and visualization are publicly available. For instance, the R package igraph (Csardi & Nepusz, 2006) or the standalone program Cytoscape (Shannon *et al.*, 2003) can be used to analyse and visualize graphs. Cytoscape also provides easy integration of biological databases such as Gene Ontology (Ashburner *et al.*, 2000), Reactome (Croft *et al.*, 2014), the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) or BioGRID (Chatr-Aryamontri *et al.*, 2013) by third-party apps. Several methods were developed to identify modules of nodes which are jointly affected by the condition of interest. Two publicly available examples are the Cytoscape plugin jActiveModules (Ideker *et al.*, 2002) and the R package BioNet (Beisser *et al.*, 2010).

Here, we present a selection of current methods to construct and analyse biological networks as an approach to systems biology and their impact on aging research.

Enrichment and network topology analysis in predefined networks

A popular approach to put the results of an association study in a systems biology context is projecting the variables of interest – such as age-related genes, proteins or metabolites – onto known biological (reference) networks. The neighbourhood of these target variables and their topological properties can then be assessed using the experimentally predefined PPI, GRN or metabolic networks. Instead of interpreting individual entities separately, *a priori* knowledge about their interactions and common functions can be used to identify modules that are jointly affected by the condition of interest.

Several databases offer a collection of experimentally identified interactions that can be used as predefined reference networks for enrichment and topology. In case of PPI, the Human Protein Reference Database provides more than 40 000 PPIs (Keshava Prasad *et al.*, 2009), the Database of Interacting Proteins more than 7000 interactions (Xenarios *et al.*, 2002) and the MIPS mammalian protein–protein database roughly 1000 hand-curated interactions of human proteins (Pagel *et al.*, 2005). GRN are provided by the ChIPBase (Yang *et al.*, 2013), which contains six million transcription factor binding sites from >300 experiments. Metabolic reactions are amongst others provided by KEGG.

Enrichment analysis is a convenient way to incorporate existing knowledge from biological reference networks without analysing graph topology directly. Therefore, predefined (functional) modules within the reference networks are used to test overrepresentation of associated genes, proteins or metabolites in these groups. When investigating genes, researchers usually use Gene Ontology to group genes based on *biological processes*, *molecular functions* or *subcellular localization*. For metabolites, the KEGG and Reactome databases provide curated information about biochemical pathways. The R packages GSEABase, GAGE (Luo *et al.*, 2009) and the webservice MSEA (Xia & Wishart, 2010)

are just some of many available implementations and variations in the original gene set enrichment analysis (Subramanian *et al.*, 2005) algorithm.

In aging research, enrichment analysis unveiled an overexpression of genes involved in immune response, lysosome and glycoproteins and an underexpression of mitochondrial- and oxidative phosphorylation-related genes in old people compared with young (de Magalhães *et al.*, 2009). In human brain tissue, oxidative stress/DNA repair and inflammation-related genes were shown to be enriched in the set of differentially expressed genes between young and old individuals (Lu *et al.*, 2004). Enrichment analysis facilitates the identification of pathways that are important for the aging process. It thus helps to make sense out of the individual associations and find biological interpretations for the observed molecular changes.

To become independent of predefined module annotation and to enable more detailed network analysis, the variables of interest can also be mapped directly on the known PPI, GRN or metabolism networks. Modules can then be identified dynamically based on the measured data. Moreover, additional topological properties of the variables of interest can be assessed.

Studying human PPI networks revealed that genes that are associated with aging by homology have higher node degrees and higher betweenness centrality compared with other genes (Bell *et al.*, 2009). Furthermore, aging-related genes are not spread throughout the interactome, but cluster in few tightly connected modules. These modules were enriched in DNA damage repair and stress response genes (Kriete *et al.*, 2011). The high connectivity of aging genes was used by Tacutu *et al.* (2012) to select neighbours of longevity-related genes in a PPI network as longevity-gene candidates. Subsequent experiments in *C. elegans* revealed 30 new longevity-associated genes, proving the potential of network biology for candidate gene selection. Using a modified PPI network, Wang *et al.* (2009) showed a tight connection of the genetic causes of aging and disease. These results indicate that aging does not occur due to random errors but is an organized process. Another PPI-based approach to data integration was developed by West *et al.* (2013). They incorporated epigenomic data by assigning DNA methylation sites to each protein in the graph and then identifying modules of differentially methylated genes/proteins in the resulting network. By doing so, they avoided predefined gene sets as used by enrichment analysis. The analysis revealed three differentially methylated modules, which were replicated across several tissues. Two of them contained mainly transcription regulating genes, while the third one contained genes related to stem cell differentiation.

A drawback of experimentally derived PPI or GRN is that such methods detect up to 50% false positives while many true interactions are missed (Huang & Bader, 2009; Marbach *et al.*, 2012). Even more importantly, those reference networks completely ignore the tempo-spatial properties of the interactions. This restricts results to already observed, possibly inactive interactions. One method to overcome the static nature of PPI networks are Negative-Positive (NP) networks (Xia *et al.*, 2006). These integrate the PPI network with transcriptomics data by restricting it to edges between (anti-)correlated proteins/genes. Therefore, only those interactions (=edges) that are active under the observed condition are further analysed. Xue *et al.* (2007) applied this method to the previously mentioned data set of brain gene expression and unveiled two anticorrelated modules containing cell proliferation- and cell differentiation-related proteins. Two other modules consisting of protein processing and immunity-related genes, respectively, were found to be slightly correlated with the cell proliferation module. A recent study went one step further and restricted a PPI network to highly expressed

genes in different stages of aging for each sample separately, thus generating a set of *dynamic binding networks* instead of a single network. Even though the global properties of all those graphs were very similar, the centrality of several genes correlated with age (Faisal & Milenković, 2014).

Incorporating biological networks to analyse aging-related changes showed the tight connection of aging and disease on a molecular level. Furthermore, it has been shown that aging affects central genes, which are important for the network integrity (Bell *et al.*, 2009). While network-based enrichment and analysis using PPI networks is common for genetic and transcriptomics data, it has not been applied to aging studies using metabolomics data. This could be a promising approach to systematically identify metabolic pathways jointly affected by the aging process.

Analysis of data-derived networks

Despite their successful applications, all approaches presented so far rely on predefined, static networks. To overcome the limitations of such networks, inferring networks directly from the measured data is the next step.

Weighted gene co-expression network analysis

The weighted gene co-expression network analysis (WGCNA) (Zhang & Horvath, 2005) infers gene-gene interaction networks directly from transcriptomics data. Miller *et al.* (2008) applied this method to the previously mentioned gene expression data set of 30 human frontal cortex samples at different ages and then compared the results with a network derived from an AD transcriptomics study. It revealed significant overlap between healthy aging and AD, suggesting that there might be a shared molecular basis for both processes. Three AD network modules overlapped with aging network modules, containing mostly synapses-, transport- and transcriptional regulation-related genes.

Gaussian graphical models

Despite the successful application of WGCNA on transcriptomics data, Krumsiek *et al.* (2011) showed that ordinary correlations are not suitable to analyse metabolomics data from large cohort studies. They analysed metabolite concentrations of >1000 samples and found that more than half of all pairs of 151 metabolites correlated significantly, even when using a restrictive Bonferroni correction at an alpha level of 0.01. This is largely due to indirect associations, which cannot be distinguished from direct associations by the Pearson correlation coefficient. Graphical models (GMs), also known as conditional independence graphs, were proposed to overcome this problem and infer biological meaningful networks from metabolomics (Steuer, 2006; Krumsiek *et al.*, 2011) as well as other omics data (de la Fuente *et al.*, 2004; Yuan *et al.*, 2011; Mangin *et al.*, 2012). GMs are probabilistic models where an edge between two variables illustrates their *conditional dependence* given all other variables in the model. Implicitly, the absence of an edge represents the conditional independence of the according variables. Several algorithms to infer GMs from purely binary data are publicly available as R packages (Wainwright *et al.*, 2006; Höfling & Tibshirani, 2009; Guo *et al.*, 2010; Ravikumar *et al.*, 2010). Their counterparts for purely continuous data are Gaussian graphical models (GGMs), which use partial correlations to infer graphs. A partial correlation of two variables X and Y conditioned on a set of variables Z quantifies the portion of the correlation between X and Y which cannot be attributed to Z. Several algorithms exist to infer GGMs (d'Aspremont *et al.*, 2006; Meinshausen & Bühlmann, 2006; Yuan & Lin, 2007; Friedman *et al.*,

Table 1 Overview over system biology methods and their application in aging

Method	Prerequisites	Applies to	Availability	Application
Enrichment Analysis	Module definition (e.g. gene sets from Gene Ontology)	Genomics Transcriptomics Proteomics Metabolomics	Several R packages (e.g. GSEABase, GAGE, MSEA), online tools DAVID or Enrichr	Lu <i>et al.</i> (2004), de Magalhães <i>et al.</i> (2009)
Network Mapping	Predefined network, such as protein–protein interaction (PPI) networks, gene regulatory network (GRN) or metabolic network	Any omics data	R package igraph, Cytoscape with various plugins	Wang <i>et al.</i> (2009), Bell <i>et al.</i> (2009), West <i>et al.</i> (2013), Faisal & Milenković (2014)
NP Networks	PPI Network	Transcriptomics	–	Xue <i>et al.</i> (2007)
Weighted Gene Co-Expression Network Analysis (WGCNA)	–	Transcriptomics (and possibly other continuous data)	R package WGCNA	Miller <i>et al.</i> (2008)
Gaussian graphical models (GGMs)	–	Any multivariate Gaussian distributed data	Several R packages (e.g. ggm or glasso)	Applied to metabolomics data by Krumsiek <i>et al.</i> (2011)
Mixed graphical models (MGMs)	–	Binary, continuous and mixed data	–	–
Bayesian Networks	–	Binary, continuous and mixed data	Several R packages (e.g. bnlearn, gRain, abn, deal)	Applied to transcriptomics data by Friedman <i>et al.</i> (2000)

2008; Mazumder & Hastie, 2012). Several of them, such as the well-established graphical lasso (Friedman *et al.*, 2008; Mazumder & Hastie, 2012), use regularization to further reduce the number of edges in the graph. This allows researchers to concentrate on fewer high-confidence interactions.

Gaussian graphical models can reconstruct biological pathways from metabolomics and transcriptomics data, but have not yet been applied in aging research. However, their application could help reduce the ‘overabundance’ of results to fewer, meaningful associations. The major drawback of GGMs is that they can only be used for pure Gaussian or pure binary data. Shin *et al.* (2014) overcame this problem by first constructing a GGM from metabolite concentrations and then adding gene variants as nodes and connecting them with associated metabolites. The resulting network illustrates the genetic control of the metabolism in an intuitive way. However, it is no longer a GM, and edges do not indicate conditional independence any more.

Mixed graphical models

Recent developments allow the integration of different types of data while maintaining the favourable properties of GGMs, namely mixed graphical models (MGMs) (Tur & Castelo, 2012; Chen *et al.*, 2013; Fellinghauer *et al.*, 2013; Lee & Hastie, 2015). Fellinghauer *et al.* (2013) proposed a very flexible algorithm based on stability selection (Meinshausen & Bühlmann, 2010). It makes use of established methods such as random forests or regression models to rank interactions between variables of different types. Thus, it can handle many different data types such as disease states, metabolite levels and gene variants. Due to the usage of stability selection, it has an intrinsic error control. MGMs provide a powerful tool for multivariate analyses of high-dimensional data, but have not been applied in biological research, yet. Their application could shed light on the complex relationship between aging and disease.

Gaussian graphical models as well as MGMs are undirected models. Therefore, neither of them can be used to infer causal direction. In epidemiological research, Mendelian randomization is a common approach to infer causality from observational data. It takes advantage of the invariability of gene variants to separate the study population in groups, thus mimicking a randomized controlled trial (for further details,

see Brion *et al.*, 2014). Mendelian randomization can be used to further investigate edges of interest that were previously identified by GGMs. However, it relies on stable associations with genetic variants and assumes that this genetic variant is not related to any other potential confounding factor. Due to these restrictions, it is not suitable to infer large-scale networks.

Bayesian networks

Another approach that allows inferring causality from observational data under certain assumptions is based on Bayesian networks (BNs). Similar to GGMs, BNs are probabilistic models in which edges represent the conditional independence between variables. However, BNs are DAGs, thus distinguishing between an influence of X on Y and the influence of Y on X. In return, the acyclicity of the causal graph is an assumption which might not hold true for biological networks. The application of BNs on high-throughput transcriptomics data by Friedman *et al.* (2000) demonstrated the potential of this method to extract biological meaningful associations without prior knowledge. Several different methods are available to estimate the structure of BNs from binary, continuous and even mixed data such as the R packages bnlearn (Scutari, 2010) (Table 1).

The methods presented here are just a selection of the available methods for graph inference. Several other methods such as Boolean networks (Shmulevich *et al.*, 2002) or differential equation systems (Chen *et al.*, 1999; Lorenz *et al.*, 2009) are commonly used for modelling biological networks.

The development of new techniques facilitates graph inference from high-dimensional data, and the presented studies illustrate their usefulness in biological research. However, most graph inference methods rely on large sample sizes and usually more samples than variables are needed. When analysing omics data, particularly genomics or transcriptomics, this is often not feasible and it is referred to as the $n \ll p$ problem. Another common problem is overfitting of models due to the high number of parameters. Some techniques such as regularization have been proposed to relax these constraints and reduce overfitting. Nevertheless, stringent cross-validation and replication in independent cohorts should be employed to avoid spurious results. Finally, many high-throughput methods suffer from considerable technical variation and strong batch effects. Researchers should carefully normalize all mea-

measurements according to current standards before integrating different data sets.

Model biological systems

The ultimate goal of systems biology is not only the qualitative exploration, but the quantitative modelling of the organism, facilitating *in silico* experiments, hypotheses generation and predictions.

The first – and so far only – attempt to model a whole organism was conducted by Karr *et al.* (2012). They created a model of a mycoplasma genitalium cell simulating cell cycle and predicting metabolite concentrations. However, the model is far from perfect (Freddolino & Tavazoie, 2012) and too primitive to be adapted to more complex organisms. Currently, modelling eukaryotic cells or even whole organisms is not feasible. Also, processes like aging are too complex to be entirely modelled. However, some effort has been undertaken to create network representations of smaller subsystems as well as certain aspects of the aging process. For instance, Gillespie *et al.* (2004) simulated aging of yeast based on the accumulation of extrachromosomal ribosomal DNA circles. Also, Oda & Kitano (2006) summarized results from several hundred studies to create a model of the Toll-like receptor (TLR) signalling network. The same group also created a similar model for epidermal growth factor receptor signalling (Oda *et al.*, 2005). Both studies revealed a bowtie-like global structure with one important key regulator. However, both networks are only qualitative descriptions without kinetic parameters. Thus, they cannot be used for computer simulations.

Other groups concentrated on even smaller subsystems to facilitate quantitative modelling. One study investigated the influence of increased cortisol levels on hippocampus activity (McAuley *et al.*, 2009). A quantitative model was created to simulate the decline in hippocampal output with age and the acceleration of this process due to acute and chronic increases in cortisol levels. Simulations using ordinary differential equations suggested that chronic increase in cortisol levels leads to faster decline in hippocampal output than acute bursts, but could be treated more efficiently. Sozou & Kirkwood (2001) modelled cell senescence based on telomere shortening and oxidative stress. The same group also described the influence of chaperones and accumulation of misfolded proteins on aging (Proctor *et al.*, 2005). Other groups investigated various further aspects of the aging process, such as mitochondrial fusion and fission events and accumulation of defective mitochondria (Kowald *et al.*, 2005; Figge *et al.*, 2012), incomplete replication of epigenetic information (Przybilla *et al.*, 2014) and age-related alterations in the lipid metabolism (McAuley & Mooney, 2015). Adjusting the kinetics of such models to correspond to experimental observations allows to come up with plausible hypotheses about the causes of aging.

In contrast to earlier presented networks, which inferred large-scale networks from data (top-down approach), these approaches model small subsystems in high details based on expert *a priori* knowledge (bottom-up approach). Such bottom-up models allow mechanistic insights into the processes of aging that cannot be generated by individual association studies. Moreover, they facilitate the development of new hypothesis and testing the plausibility of current hypothesis.

Conclusions and challenges

The major recent advances of omics technologies are now enabling the simultaneous measurement of millions of biochemical entities. Association studies have revealed many associations of omics data with aging and age-related diseases. After decades of reductionist studies, network analysis and integrated omics data analysis have begun to target the

aging process at a systems level. As a result, some studies take into account also the interaction effects between variables. However, given the complexity of aging, new methods are needed to further unveil the multiple interactions.

Systems biology already provides such methods, but their application on real biological problems lags behind. For example, GGMs have been adapted to mixed data types and could readily be applied in aging research. Also, several studies developed models of processes that contribute to aging. These provide detailed knowledge about important components of the aging process and their interactions. Building on these results, future studies should aim to integrate these different parts to gain a more systems-level understanding of aging.

However, in many cases, the available data limit the possibilities. Problems such as incomplete data, asynchronous experiments, strong batch effects and insufficient sample sizes have to be dealt with. Another issue is the limited availability of multi-omics data sets, which complicates replication of results in this field. A variety of different methods, protocols and platforms further hampers reproducible results. As replication of results is crucial to prevent spurious results and validation, methods like splitting the available data into discovery and replication sets should be considered more often.

Despite these obstacles, there are several large population studies in existence with multi-omics data available which could be explored using systems biology approaches. For instance, the GTEx project aims to collect gene expression and methylation data from multitissue samples (The GTEx Consortium 2013). Simultaneously, the development of new methods should help to analyse real, partially incomplete data sets and facilitate analysis of multitissue and multi-organ data, thus enabling the investigation of real systems-level effects. Addressing these problems and developing integrated models of aging should improve our understanding of the aging process, thus allowing the development of strategies to improve health in old age.

Funding

This work was supported by the EU Framework Programme 7 small-scale focused research collaborative project EurHEALTHAging [277849]; TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme [FP7/2007-2013]. The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TDS is a NIHR senior research fellow.

Conflict of interest

None declared.

References

- Adamczyk B, Tharmalingam T, Rudd PM (2012) Glycans as cancer biomarkers. *Biochim. Biophys. Acta* **1820**, 1347–1353.
- Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. *Nature* **406**, 378–382.
- Alegria-Torres JA, Baccarelli A, Bollati V (2011) Epigenetics and lifestyle. *Epigenomics* **3**, 267–277.
- Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* **5**, 18.
- Alzheimer's Society (2014) Demography. Available at: <http://www.alzheimers.org.uk>.

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29.
- d'Aspremont A, Banerjee O, El Ghaoui L (2006) First-order methods for sparse covariance selection. *J. Mach. Learn. Res.* **10**, 883–906.
- Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, Santoro A, Franceschi C, Garagnani P (2014) Present and future of anti-ageing epigenetic diets. *Mech. Ageing Dev.* **136–137**, 101–115.
- Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113.
- Baylis D, Ntani G, Edwards MH, Syddall HE, Bartlett DB, Dennison EM, Martin-Ruiz C, von Zglinicki T, Kuh D, Lord JM, Aihie Sayer A, Cooper C (2014) Inflammation, telomere length, and grip strength: a 10-year longitudinal study. *Calcif. Tissue Int.* **95**, 54–63.
- Beisser D, Klau GW, Dandekar T, Müller T, Dittrich MT (2010) BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics* **26**, 1129–1130.
- Bell R, Hubbard A, Chettier R, Chen D, Miller JP, Kapahi P, Tarnopolsky M, Sahasrabudhe S, Melov S, Hughes RE (2009) A human protein interaction network shows conservation of aging processes between human and invertebrate species. *PLoS Genet.* **5**, e1000414.
- Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin S-Y, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, Dermizakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* **8**, e1002629.
- Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, Beach T, Miller C, Troncoso J, Trojanowski JQ, Zielke HR, Cotman CW (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl Acad. Sci. U. S. A.* **105**, 15605–15610.
- Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, Nikkila J, Monti D, Satokari R, Franceschi C, Brigidi P, De Vos W (2010) Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**, e10667.
- Biagi E, Candela M, Fairweather-Tait S, Franceschi C, Brigidi P (2012) Aging of the human metagenome: the microbial counterpart. *Age (Dordt)*. **34**, 247–267.
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* **88**, 450–457.
- Brion MA, Benyamin B, Visscher PM, Smith GD (2014) Beyond the single SNP: emerging developments in mendelian randomization in the 'omics' era. *Curr. Epidemiol. Reports* **1**, 228–236.
- Cassman M (2005) Barriers to progress in systems biology. *Nature* **438**, 1079.
- Cevenini E, Invidia L, Lescai F, Salvioli S, Tieri P, Castellani G, Franceschi C (2008) Human models of aging and longevity. *Expert Opin. Biol. Ther.* **8**, 1393–1405.
- Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* **41**, D816–D823.
- Chen T, He HL, Church GM (1999) Modeling gene expression with differential equations. *Pac. Symp. Biocomput.* **4**, 29–40.
- Chen S, Witten D, Shojai A (2013) Selection and Estimation for Mixed Graphical Models. Available at: <http://arxiv.org/abs/1311.0085> [Accessed January 7, 2014].
- Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, Marchesi JR, Falush D, Dinan T, Fitzgerald G, Stanton C, van Sinderen D, O'Connor M, Harnedy N, O'Connor K, Henry C, O'Mahony D, Fitzgerald AP, Shanahan F, Twomey C, Hill C, Ross RP, O'Toole PW (2011) Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl Acad. Sci. U. S. A.* **108**(Suppl), 4586–4591.
- Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HMB, Coakley M, Lakshminarayanan B, O'Sullivan O, Fitzgerald GF, Deane J, O'Connor M, Harnedy N, O'Connor K, O'Mahony D, van Sinderen D, Wallace M, Brennan L, Stanton C, Marchesi JR, Fitzgerald AP, Shanahan F, Hill C, Ross RP, O'Toole PW (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184.
- Craig T, Smelick C, Tacutu R, Wuttke D, Wood SH, Stanley H, Janssens G, Savitskaya E, Moskalev A, Arking R, de Magalhães JP (2014) The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource. *Nucleic Acids Res.* **43**, D873–D878.
- Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477.
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Sy.* **1695**, Available at: <http://igraph.org>.
- Dang W, Steffen KK, Perry R, Dorsey JA, Johnson FB, Shilatfard A, Kaerberlein M, Kennedy BK, Berger SL (2009) Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802–807.
- Davies G, Harris SE, Reynolds CA, Payton A, Knight HM, Liewald DC, Lopez LM, Luciano M, Gow AJ, Corley J, Henderson R, Murray C, Pattie A, Fox HC, Redmond P, Lutz MW, Chiba-Falek O, Linnertz C, Saith S, Haggarty P, McNeill G, Ke X, Ollier W, Horan M, Roses AD, Ponting CP, Porteous DJ, Tenesa A, Pickles A, Starr JM, Whalley LJ, Pedersen NL, Pendleton N, Visscher PM, Deary IJ (2014) A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Mol. Psychiatry* **19**, 76–87.
- Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HED, Lakenberg N, van den Akker EB, Passtoors WM, Tiemeier H, van Heemst D, de Craen AJ, Rivadeneira F, de Geus EJ, Perola M, van der Ouderaa FJ, Gunn DA, Boomsma DI, Uitterlinden AG, Christensen K, van Duijn CM, Heijmans BT, Houwing-Duistermaat JJ, Westendorp RGJ, Slagboom PE (2011) Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Ageing Cell* **10**, 686–698.
- Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–929.
- Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. *Oncogene* **21**, 5400–5413.
- Faisal FE, Milenković T (2014) Dynamic networks reveal key players in aging. *Bioinformatics* **30**, 1721–1729.
- Fellinghauer B, Bühlmann P, Ryffel M, von Rhein M, Reinhardt JD (2013) Stable graphical model estimation with Random Forests for discrete, continuous, and mixed variables. *Comput. Stat. Data Anal.* **64**, 132–152.
- Figge MT, Reichert AS, Meyer-Hermann M, Osiewicz HD (2012) Deceleration of fusion-fission cycles improves mitochondrial quality control during aging. *PLoS Comput. Biol.* **8**, e1002576.
- Flachsbart F, Caliebe A, Kleindorp R, Blanché H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl Acad. Sci. U. S. A.* **106**, 2700–2705.
- Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815.
- Freddolino PL, Tavazoie S (2012) The dawn of virtual cell biology. *Cell* **150**, 248–250.
- Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620.
- Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441.
- de la Fuente A, Bing N, Hoeschele I, Mendes P (2004) Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **20**, 3565–3574.
- Fuster MM, Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer* **5**, 526–542.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174.
- Gelfi C, Viganò A, Ripamonti M, Pontoglio E, Nica AC, Meglio P, Nestle FO, Ryten M, Durbin R, McCarthy MI, Deloukas P, Dermizakis ET, Weale ME, Bataille V, Spector TD (2013) Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14**, R75.
- Goh K, Cusick ME, Valle D, Vidal M, Barabási A-L (2007) The human disease network. *Proc. Natl Acad. Sci. U. S. A.* **104**, 8685–8690.
- Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhan R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE (2014) Human genetics shape the gut microbiome. *Cell* **159**, 789–799.

- Gornik O, Wagner J, Pucić M, Knezević A, Redzic I, Lauc G (2009) Stability of N-glycan profiles in human plasma. *Glycobiology* **19**, 1547–1553.
- Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* **466**, 383–387.
- Grice EA, Segre JA (2012) The human microbiome: our second genome. *Annu. Rev. Genomics Hum. Genet.* **13**, 151–170.
- Guigoz Y, Doré J, Schiffrin EJ (2008) The inflammatory status of old age can be nurtured from the intestinal environment. *Curr. Opin. Clin. Nutr. Metab. Care* **11**, 13–20.
- Guo J, Levina E, Michailidis G, Zhu J (2010) Joint Structure Estimation for Categorical Markov Networks.
- Ha N-T, Freytag S, Bickeboeller H (2014) Coverage and efficiency in current SNP chips. *Eur. J. Hum. Genet.* **22**, 1124–1130.
- Hammond CJ, Duncan DD, Snieder H, de Lange M, West SK, Spector TD, Gilbert CE (2001) The heritability of age-related cortical cataract: the twin eye study. *Invest. Ophthalmol. Vis. Sci.* **42**, 601–605.
- Harman D (1988) The aging process. *Basic Life Sci.* **49**, 1057–1065.
- Harman D (2001) Aging: overview. *Ann. N. Y. Acad. Sci.* **928**, 1–21.
- Höfling H, Tibshirani R (2009) Estimation of sparse binary pairwise markov networks using pseudo-likelihoods. *J. Mach. Learn. Res.* **10**, 883–906.
- Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115.
- Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. *Nat. Rev. Genet.* **11**, 855–866.
- Houseman E, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86.
- Huang H, Badier JS (2009) Precision and recall estimates for two-hybrid screens. *Bioinformatics* **25**, 372–378.
- Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.
- Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**(Suppl 1), S233–S240.
- Illumina (2011) Infinium HumanMethylation450 BeadChip Kit. Available at: http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn [Accessed June 11, 2014].
- Ishimori ML, Altman RD, Cohen MJ, Cui J, Guo X, Rotter JJ, Weisman MH (2010) Heritability patterns in hand osteoarthritis: the role of osteophytes. *Arthritis. Res. Ther.* **12**, R180.
- Jackson SHD, Weale MR, Weale RA (2003) Biological age—what is it and can it be measured? *Arch. Gerontol. Geriatr.* **36**, 103–115.
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**(Suppl), 245–254.
- Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15**, R31.
- Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* **411**, 41–42.
- Johnson TE (2006) Recent results: biomarkers of aging. *Exp. Gerontol.* **41**, 1243–1246.
- Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network. *J. Biomed. Biotechnol.* **2005**, 96–103.
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.
- Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadrana S, Chaerkady R, Pandey A (2009) Human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767–D772.
- Kirkwood TB, Austad SN (2000) Why do we age? *Nature* **408**, 233–238.
- Knezević A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, Wright A, Kolcic I, O'Donoghue N, Bones J, Rudd PM, Lauc G (2009) Variability, heritability and environmental determinants of human plasma N-glycome. *J. Proteome Res.* **8**, 694–701.
- Kong HH (2011) Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends Mol. Med.* **17**, 320–328.
- Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335.
- Kowald A, Jendrach M, Pohl S, Bereiter-Hahn J, Hammerstein P (2005) On the relevance of mitochondrial fusions for the accumulation of mitochondrial deletion mutants: a modelling study. *Aging Cell* **4**, 273–283.
- Kriete A, Lechner M, Clearfield D, Bohmann D (2011) Computational systems biology of aging. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 414–428.
- Kristic J, Vuckovic F, Menni C, Klaric L, Keser T, Beceheli I, Pucic-Bakovic M, Novokmet M, Mangino M, Thaqi K, Rudan P, Novokmet N, Sarac J, Missoni S, Kolcic I, Polasek O, Rudan I, Campbell H, Hayward C, Aulchenko Y, Valdes A, Wilson JF, Gornik O, Primorac D, Zoldos V, Spector T, Lauc G (2013) Glycans are a novel biomarker of chronological and biological ages. *J. Gerontol. A Biol. Sci. Med. Sci.* **69**, 1–11.
- Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ (2011) Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* **5**, 21.
- Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, Hanson RW, Kalhan SC, Ryals JA, Milburn MV (2008) Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397.
- Lee JD, Hastie TJ (2015) Learning the structure of mixed graphical models. *J. Comput. Graph. Stat.* **24**, 230–253.
- Levine ME (2013) Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J. Gerontol. A Biol. Sci. Med. Sci.* **68**, 667–674.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 1–11.
- Lodish HF (2013) *Molecular Cell Biology*, 7th edn. New York: W.H. Freeman and Co.
- López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* **153**, 1194–1217.
- Lorenz DR, Cantor CR, Collins JJ (2009) A network biology approach to aging in yeast. *Proc. Natl Acad. Sci. U. S. A.* **106**, 1145–1150.
- Lu T, Pan Y, Kao S, Li C, Kohane I, Chan J, Yankner BA (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891.
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161.
- de Magalhães JP, Curado J, Church GM (2009) Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881.
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity (Edinb.)* **108**, 285–291.
- Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804.
- Mazumder R, Hastie T (2012) The graphical lasso: new insights and alternatives. *Electron. J. Stat.* **6**, 2125–2149.
- McAuley MT, Mooney KM (2015) Computationally modeling lipid metabolism and aging: a mini-review. *Comput. Struct. Biotechnol. J.* **13**, 38–46.
- McAuley MT, Kenny RA, Kirkwood TBL, Wilkinson DJ, Jones JLL, Miller VM (2009) A mathematical model of aging-related and cortisol induced hippocampal dysfunction. *BMC Neurosci.* **10**, 26.
- Mehra VC, Ramgolam VS, Bender JR (2005) Cytokines and cardiovascular disease. *J. Leukoc. Biol.* **78**, 805–818.
- Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.* **34**, 1436–1462.
- Meinshausen N, Bühlmann P (2010) Stability selection. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **72**, 417–473.
- Meissner A (2010) Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol.* **28**, 1079–1088.
- Menni C, Fauman E, Erte I, Perry JRB, Kastenmüller G, Shin SY, Petersen AK, Hyde C, Psatha M, Ward KJ, Yuan W, Milburn M, Palmer CNA, Frayling TM, Trimmer J, Bell JT, Gieger C, Mohny RP, Brosnan MJ, Suhre K, Soranzo N, Spector TD (2013a) Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* **62**, 4270–4276.
- Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, Lim EM, Stuckey B, Deloukas P, Mohny R, Suhre K, Spector TD, Valdes AM (2013b) Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int. J. Epidemiol.* **42**, 1111–1119.
- Menni C, Kiddle SJ, Mangino M, Vinuela A, Psatha M, Steves C, Sattler M, Buil A, Newhouse S, Nelson S, Williams S, Voyle N, Soininen H, Kloszewska I,

- Mecocci P, Tsolaki M, Vellas B, Lovestone S, Spector TD, Dobson R, Valdes AM (2015) Circulating proteomic signatures of chronological age. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 809–816.
- Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* **28**, 1410–1420.
- Mitnitski A, Song X, Rockwood K (2013) Assessing biological aging: the origin of deficit accumulation. *Biogerontology* **14**, 709–717.
- Moayyeri A, Hammond CJ, Valdes AM, Spector TD (2013) Cohort profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* **42**, 76–85.
- Mungreiphy NK, Kapoor S, Sinha R (2011) Association between BMI, blood pressure, and age: study among tangkhul naga tribal males of northeast India. *J. Anthropol.* **2011**, 1–6.
- Murabito JM, Yuan R, Lunetta KL (2012) The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. *J. Gerontol. A Biol. Sci. Med. Sci.* **67**, 470–479.
- Nakajima K, Takeoka M, Mori M, Hashimoto S, Sakurai A, Nose H, Higuchi K, Itano N, Shiohara M, Oh T, Taniguchi S (2010) Exercise effects on methylation of ASC gene. *Int. J. Sports Med.* **31**, 671–675.
- Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanché H, Junge O, Wittig M, Ellinghaus D, Flachsbarf F, Wichmann HE, Meitinger T, Nikolaus S, Franke A, Krawczak M, Lathrop M, Schreiber S (2011) A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* **132**, 324–330.
- Oda K, Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* **2**, 2006.0015.
- Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* **1**, 2005.0010.
- Oeppen J, Vaupel JW (2002) Demography. Broken limits to life expectancy. *Science* **296**, 1029–1031.
- Office for National Statistics (2014) Historic and Projected Data from the Period and Cohort Life Tables, 2012-based revised. Available at: <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcn:77-355125>.
- Orešić M, Hyötyläinen T, Herukka S-K, Sysi-Aho M, Mattila I, Seppänen-Laakso T, Julkunen V, Gopalacharyulu PV, Hallikainen M, Koikkalainen J, Kivipelto M, Helisalmi S, Lötjönen J, Soininen H (2011) Metabolome in progression to Alzheimer's disease. *Transl. Psychiatry* **1**, e57.
- Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes H-W, Ruepp A, Frishman D (2005) The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834.
- Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269.
- Petersen AK, Zeilinger S, Kastenmüller G, Werner RM, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P, Huber F, Mohny RP, Grallert H, Illig T, Adamski J, Waldenberger M, Gieger C, Suhre K (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum. Mol. Genet.* **23**, 534–545.
- Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M (2009) The NIH human microbiome project. *Genome Res.* **19**, 2317–2323.
- Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068.
- Proctor CJ, Soti C, Boys RJ, Gillespie CS, Shanley DP, Wilkinson DJ, Kirkwood TBL (2005) Modelling the actions of chaperones and their role in ageing. *Mech. Ageing Dev.* **126**, 119–131.
- Przybilla J, Rohlf T, Loeffler M, Galle J (2014) Understanding epigenetic changes in aging stem cells - a computational model approach. *Ageing Cell* **13**, 320–328.
- Pucić M, Knezević A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redžić I, Campbell H, Wright A, Hastie ND, Wilson JF, Rudan I, Wuhrer M, Rudd PM, Josić D, Lauc G (2011) High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol. Cell Proteomics* **10**, M111.010090.
- Rattan SIS (2006) Theories of biological aging: genes, proteins, and free radicals. *Free Radic. Res.* **40**, 1230–1238.
- Ravikumar P, Wainwright MJ, Lafferty JD (2010) High-dimensional Ising model selection using ℓ_1 -regularized logistic regression. *Ann. Stat.* **38**, 1287–1319.
- Reeve A, Simcox E, Turnbull D (2014) Ageing and Parkinson's disease: why is advancing age the biggest risk factor?. *Ageing Res. Rev.* **14C**, 19–30.
- Rockwood K, Mitnitski A (2007) Frailty in relation to the accumulation of deficits. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **62**, 722–727.
- Rodwell GEJ, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, Xiao W, Mindrinos M, Crane E, Segal E, Myers BD, Brooks JD, Davis RW, Higgins J, Owen AB, Kim SK (2004) A transcriptional profile of aging in the human kidney. *PLoS Biol.* **2**, e427.
- Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 313–323.
- Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim Y-G, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA (2008) HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal. Biochem.* **376**, 1–12.
- Scutari M (2010) Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* **35**, 1–22.
- Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Hoh J, Perls TT (2012) Genetic signatures of exceptional longevity in humans. *PLoS ONE* **7**, e29848.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang T-P, Walter K, Menni C, Chen L, Vasequez L, Valdes AM, Hyde CL, Wang V, Ziemek D, Roberts P, Xi L, Grundberg E, Waldenberger M, Richards JB, Mohny RP, Milburn MV, John SL, Trimmer J, Theis FJ, Overington JP, Suhre K, Brosnan MJ, Gieger C, Kastenmüller G, Spector TD, Soranzo N (2014) An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550.
- Shmulevich I, Dougherty ER, Kim S, Zhang W (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**, 261–274.
- Smith JD (2002) Apolipoproteins and aging: emerging mechanisms. *Ageing Res. Rev.* **1**, 345–365.
- Sozou PD, Kirkwood TB (2001) A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *J. Theor. Biol.* **213**, 573–586.
- Steuer R (2006) Review: on the analysis and interpretation of correlations in metabolomic data. *Brief. Bioinform.* **7**, 151–158.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. U. S. A.* **102**, 15545–15550.
- Suhre K, Meisinger C, Döring A, Altmairer E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes H-W, Hrabé de Angelis M, Wichmann H-E, Kronenberg F, Adamski J, Illig T (2010) Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS ONE* **5**, e13953.
- Swardfager W, Lancôt K, Rothenburg L, Wong A, Cappell J, Herrmann N (2010) A meta-analysis of cytokines in Alzheimer's disease. *Biol. Psychiatry* **68**, 930–941.
- Tacutu R, Shore DE, Budovsky A, de Magalhães JP, Ruvkun G, Fraifeld VE, Curran SP (2012) Prediction of *C. elegans* longevity genes by human and worm longevity networks. *PLoS ONE* **7**, 4–12.
- Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhães JP (2013) Human ageing genomic resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* **41**, D1027–D1033.
- Teicher BA, Linehan WM, Helman LJ (2012) Targeting cancer metabolism. *Clin. Cancer Res.* **18**, 5537–5545.
- The Gtex Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585.
- The Human Microbiome Project (2014a) NIH Human Microbiome Project defines normal bacterial makeup of the body. Available at: <http://www.nih.gov/news/health/jun2012/nhgri-13.htm> [Accessed June 3, 2014].
- The Human Microbiome Project (2014b) Reference Genomes Data. Available at: <http://www.hmpdacc.org/HMRGD>.
- Tur I, Castelo R (2012) Learning high-dimensional mixed graphical models with missing values. In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models* (Cano A, Gomez-Olmedo M, Nielsen TD ed.). Granada: DECSAI, University of Granada, pp. 323–330.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. *Nature* **449**, 804–810.
- Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA (2014) Design and analysis of metabolomics studies in epidemiological research: a primer on -omic

- technologies. *Am. J. Epidemiol.* **180**, 129–139. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24966222> [Accessed June 30, 2014].
- Valdes AM, Glass D, Spector TD (2013) Omics technologies and the study of human ageing. *Nat. Rev. Genet.* **14**, 601–607.
- Varki A, Cummings R, Esko J, Freeze H, Stanley P, Bertozzi C, Hart G, Etzler M (2009) *Essentials of Glycobiology*, 2nd edn. The Consortium of Glycobiology Editors, ed., Cold Spring Harbour, NY: Cold Spring Harbor Laboratory; Auflage: 0002. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK1918/?report=reader> [Accessed June 11, 2014].
- Vijg J, Suh Y (2005) Genetics of longevity and aging. *Annu. Rev. Med.* **56**, 193–212.
- Wainwright MJ, Ravikumar P, Lafferty JD (2006) High-dimensional graphical model selection using l_1 -regularized logistic regression. In *Neural Information Processing Systems*. (B. Schölkopf and J.C. Platt and T. Hoffman, ed). Vancouver: MIT Press, pp 1465–1472.
- Wang J, Zhang S, Wang Y, Chen L, Zhang X-S (2009) Disease-aging network reveals significant roles of aging genes in connecting genetic diseases. *PLoS Comput. Biol.* **5**, e1000521.
- Warming L, Hassager C, Christiansen C (2002) Changes in bone mineral density with age in men and women: a longitudinal study. *Osteoporos. Int.* **13**, 105–112.
- Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel K-H, Erbel R, Mühleisen TW, Zenke M, Brummendorf TH, Wagner W (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **15**, R24.
- Weinert BT, Timiras PS (2003) Invited review: theories of aging. *J. Appl. Physiol.* **95**, 1706–1716.
- West J, Beck S, Wang X, Teschendorff AE (2013) An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci. Rep.* **3**, 1630.
- Wieser D, Papatheodorou I, Ziehm M, Thornton JM (2011) Computational biology for ageing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 51–63.
- Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeier S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J-H, Bantscheff M, Gerstmair A, Faerber F, Kuster B (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587.
- Willcox BJ, Donlon TA, He Q, Chen R, Grove JS, Yano K, Masaki KH, Willcox DC, Rodriguez B, Curb JD (2008) FOXO3A genotype is strongly associated with human longevity. *Proc. Natl Acad. Sci. U. S. A.* **105**, 13987–13992.
- Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorn Dahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **41**, D801–D807.
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim S-M, Eisenberg D (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**, 303–305.
- Xia J, Wishart DS (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.* **38**, W71–W77.
- Xia K, Xue H, Dong D, Zhu S, Wang J, Zhang Q, Hou L, Chen H, Tao R, Huang Z, Fu Z, Chen Y-G, Han J-DJ (2006) Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. *PLoS Comput. Biol.* **2**, e145.
- Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, Hou L, Zhang Q, Zhang Y, Han J-DJ (2007) A modular network model of aging. *Mol. Syst. Biol.* **3**, 147.
- Yang J-H, Li J-H, Jiang S, Zhou H, Qu L-H (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.* **41**, D177–D187.
- Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **3**, e59.
- Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstraß K, Wichmann H-E, Peters A, Suhre K, Li Y, Adamski J, Spector TD, Illig T, Wang-Sattler R (2012) Human serum metabolic profiles are age dependent. *Aging Cell* **11**, 960–967.
- Yuan M, Lin Y (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19–35.
- Yuan Y, Li C-T, Windram O (2011) Directed partial correlation: inferring large-scale gene regulatory network through induced topology disruptions. *PLoS ONE* **6**, e16835.
- Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17.
- Zhu B, Wang X, Li L (2010) Human gut microbiome: the second genome of human body. *Protein Cell* **1**, 718–725.
- Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481.

G.2 Metabolomics profiling reveals novel markers for leukocyte telomere length

Metabolomics profiling reveals novel markers for leukocyte telomere length

Jonas Zierer^{1,2}, Gabi Kastenmüller^{1,2}, Karsten Suhre^{2,3}, Christian Gieger^{4,5,6}, Vervan Codd⁷, Pei-Chien Tsai¹, Jordana Bell¹, Annette Peters⁵, Konstantin Strauch⁸, Holger Schulz^{9,10}, Stephan Weidinger¹¹, Robert P. Mohny¹², Nilesh J. Samani^{7,13}, Tim Spector¹, Massimo Mangino^{1,14*}, and Cristina Menni^{1*}

¹ Department of Twin Research & Genetic Epidemiology, King's College London, London, UK

² Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany

³ Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Doha, Qatar

⁴ Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

⁵ Institute of Epidemiologie II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

⁶ German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany

⁷ Department of Cardiovascular Sciences, University of Leicester, Leicester, UK

⁸ Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

⁹ Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany

¹⁰ Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich, Germany

¹¹ Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

¹² Metabolon, Inc. Durham, NC 27713, USA

¹³ National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK

¹⁴ National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, UK

* Co-senior author

Key words: telomere length, biological aging, metabolomics, glutathione, oxidative stress

Received: 10/12/15; **Accepted:** 01/17/16; **Published:** 01/20/16

Correspondence to: Cristina Menni, PhD; **E-mail:** cristina.menni@kcl.ac.uk

Copyright: Zierer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Abstract: Leukocyte telomere length (LTL) is considered one of the most predictive markers of biological aging. The aim of this study was to identify novel pathways regulating LTL using a metabolomics approach. To this end, we tested associations between 280 blood metabolites and LTL in 3511 females from TwinsUK and replicated our results in the KORA cohort. We furthermore tested significant metabolites for associations with several aging-related phenotypes, gene expression markers and epigenetic markers to investigate potential underlying pathways. Five metabolites were associated with LTL: Two lysolipids, 1-stearoylglycerophosphoinositol ($P=1.6\times 10^{-5}$) and 1-palmitoylglycerophosphoinositol ($P=1.6\times 10^{-5}$), were found to be negatively associated with LTL and positively associated with phospholipase A2 expression levels suggesting an involvement of fatty acid metabolism and particularly membrane composition in biological aging. Moreover, two gamma-

glutamyl amino acids, gamma-glutamyltyrosine ($P=2.5 \times 10^{-6}$) and gamma-glutamylphenylalanine ($P=1.7 \times 10^{-5}$), were negatively correlated with LTL. Both are products of the glutathione cycle and markers for increased oxidative stress. Metabolites were also correlated with functional measures of aging, i.e. higher blood pressure and HDL cholesterol levels and poorer lung, liver and kidney function. Our results suggest an involvement of altered fatty acid metabolism and increased oxidative stress in human biological aging, reflected by LTL and age-related phenotypes of vital organ systems.

INTRODUCTION

Telomeres are repetitive DNA sequences located at the end of each chromatid. Several proteins, such as the telomere repeat-binding factor (TRF) 1 and 2, bind specifically to this area forming large nucleoprotein complexes, the t-loops [1]. These structures protect the DNA from degradation and end-to-end fusion. Telomeres shorten with each cell cycle, due to the inability of the DNA polymerase to replicate the end of the lagging strand. Thus, the shortening of telomeres has been proposed as a “mitotic clock” which limits the replicative life span of cells and causes cell senescence [2]. In fact, leucocyte telomere length (LTL) has been associated not only with chronological age [3] but also many aging-related diseases, such as Alzheimer’s Disease (AD) [4,5], cardiovascular disease [6,7] and cancer [8,9]. Furthermore, LTL was found to predict mortality [10,11] and longevity [12]. Thus, it was suggested as potential biomarker of biological aging [13].

Genome-wide association studies have until now identified ten genes associated with LTL [14,15,16,17]. Most of these genes physically interact with telomeres; however, how the shortening of telomeres affects an individual’s health is still not fully understood.

Recent developments in the field of metabolomics allow for the high-throughput measurement of an extensive set of low-molecular-weight molecules (metabolites) [18]. Changes in metabolite concentrations reflect physiological functions and can indicate early stages of diseases [19]. Recently, a study on LTL revealed strong associations with blood biomarkers in a cohort of American Indians [20]. However, the study was small ($n=423$) and lacked independent replication.

In this study, we assess to which extent metabolic profiles are correlated with LTL in a large population study ($n=3511$, females only) from the UK using a non-targeted metabolomics platform. We replicate our results in an independent cohort from Germany ($n=904$). Furthermore, we examine the relationship of the LTL-associated metabolites with aging-related phenotypes as well as gene expression and methylation markers in order to gain insights in the mechanisms of biological aging.

RESULTS

The demographic characteristics of the study populations are presented in Table 1. We analyzed the associations between 280 fasting blood metabolites and LTL in 3511 women from the TwinsUK cohort (see supplemental Table 1).

Table 1. Population Characteristics

	TwinsUK	KORA
<i>N</i>	3511	904
<i>Age (yrs)</i>	53.6 ± 13.6	60.5 ± 8.8
<i>MZ:DZ:Singletons*</i>	1654:1360:497	0:0:904
<i>TL</i>	3.72 ± 0.67	1.85 ± 0.31
<i>BMI (kg/m²)</i>	26.21 ± 5.14	27.87 ± 5.25
<i>FEV1 (l)</i>	2.60 ± 0.61	2.79 ± 0.50
<i>HDL (mmol/L)</i>	1.71 ± 0.48	
<i>DBP (mm Hg)</i>	78.01 ± 10.68	
<i>SBP (mm Hg)</i>	126.71 ± 18.20	
<i>ALAT (IU/L)</i>	27.63 ± 17.07	
<i>GGT (U/L)</i>	28.36 ± 25.44	
<i>eGFR (mL/min/1.73m²)</i>	83.78 ± 17.07	
<i>smoking (non:ex:current)</i>	1905:1134:447	

*MZ=monozygotic, DZ=dizygotic

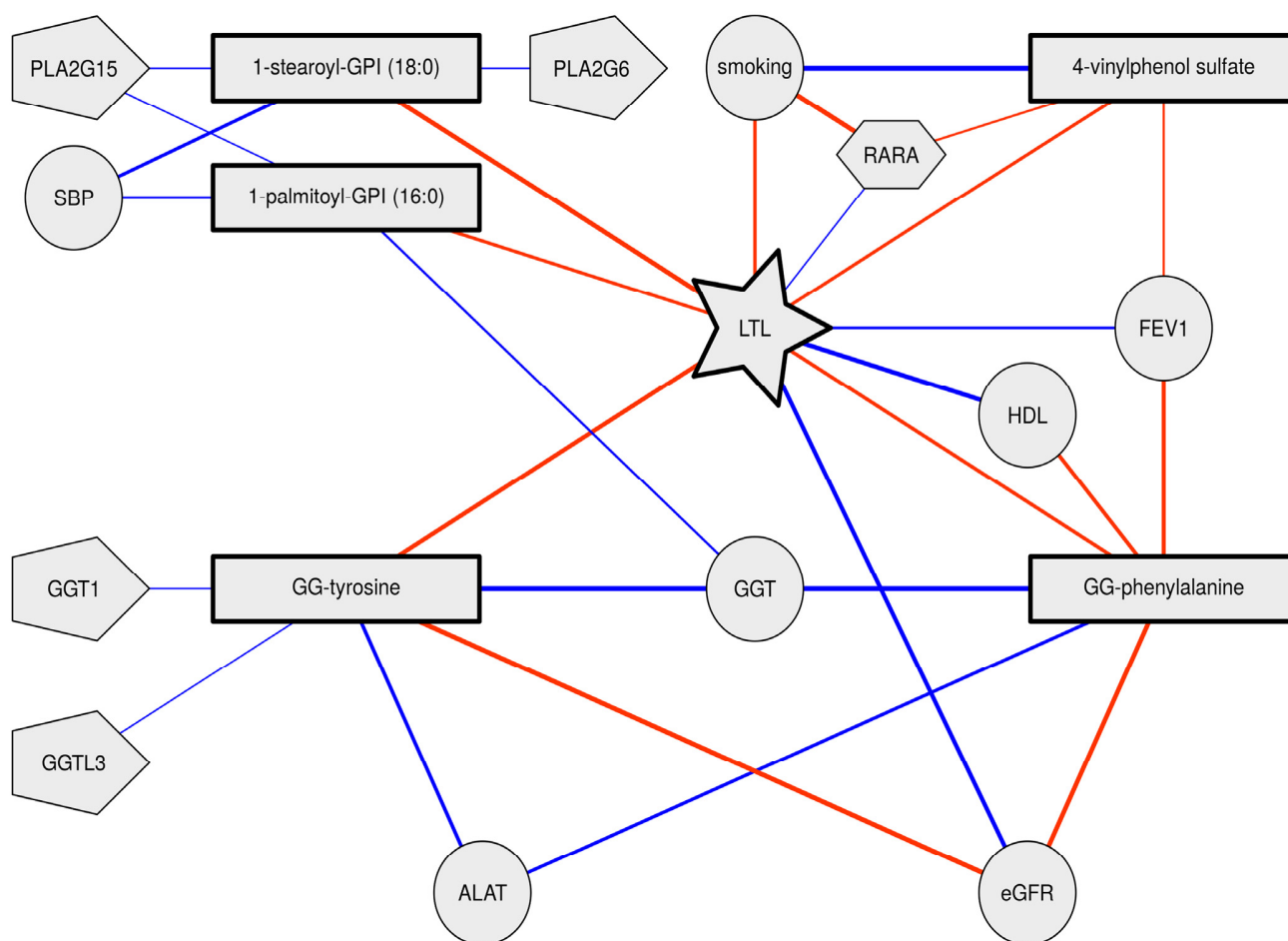


Figure 1. Telomere length, metabolite and phenotype interrelationships. Nodes represent variables where rectangles represent metabolites, circles represent phenotypes, pentagons represent expression levels and hexagons represent DNA methylation levels. Links between nodes represent significant correlations (red negative, blue positive). Thicker edges indicate stronger correlations.

Table 2. Metabolites significantly associated with LTL.

		TwinsUK		KORA	Meta	
Metabolite	PW	beta [95%CI]	p	beta [95%CI]	beta [95%CI]	p
gamma-glutamyltyrosine	Peptide	-0.09 [-0.12;-0.05]	3.41×10 ⁻⁶	-0.05 [-0.12;0.02]	-0.08 [-0.11;-0.05]	2.51×10 ⁻⁶
1-stearoylglycero-phosphoinositol	Lipid	-0.09 [-0.13;-0.05]	1.36×10 ⁻⁶	-0.00 [-0.07;0.07]	-0.07 [-0.10;-0.04]	1.60×10 ⁻⁵
1-palmitoylglycero-phosphoinositol	Lipid	-0.08 [-0.13;-0.04]	7.36×10 ⁻⁵	-0.07 [-0.14;0.01]	-0.08 [-0.12;-0.04]	1.64×10 ⁻⁵
gamma-glutamyl-phenylalanine	Peptide	-0.08 [-0.12;-0.04]	2.72×10 ⁻⁵	-0.04 [-0.11;0.02]	-0.07 [-0.10;-0.04]	1.68×10 ⁻⁵
4-vinylphenol sulfate	Xenobiotic	-0.08 [-0.12;-0.04]	7.41×10 ⁻⁵	-0.03 [-0.10;0.05]	-0.07 [-0.10;-0.03]	1.41×10 ⁻⁴

We found two lipids (1-stearoylglycerophosphoinositol: Beta [95%CI] =-0.07 [-0.10;-0.04] change in metabolite z-score per change in LTL z-score, $P=1.6\times10^{-5}$ and 1-palmitoylglycerophosphoinositol: Beta [95%CI] =-0.08 [-0.12;-0.04], $P=1.6\times10^{-5}$), two gamma-glutamyl-amino acids (gamma-glutamyltyrosine: Beta [95%CI] =-0.08 [-0.11;-0.05], $P=2.5\times10^{-6}$ and gamma-glutamylphenylalanine: Beta [95%CI] =-0.07 [-0.10;-0.04], $P=1.7\times10^{-5}$), and one xenobiotic (4-vinylphenol sulfate: Beta [95%CI] =-0.07 [-0.10;-0.03], $P=1.4\times10^{-4}$) to be negatively associated with LTL after adjustment for potential confounding factors and after correcting the results for multiple testing (Table 2, Supplemental Figure 1). All five metabolites showed the same effects with similar effect sizes in 904 female individuals from the KORA F4 study, even though they did not reach significance level. All metabolites remained Bonferroni-significant ($P<1.8\times10^{-4}$) after meta-analysis.

Three multivariate Lasso models were fitted to predict LTL: The first using clinical variables only (age, BMI), the second using the five identified metabolites only, and the third using both clinical variables and metabolites. The model based on metabolites alone could not achieve

the performance of the model based on clinical variables alone, however, combining clinical variables with metabolites significantly improved the prediction in the combined model (Figure 2). In the combined model, 1-stearoylglycerophosphoinositol was the strongest predictor followed by the 4-vinylphenol sulfate. All five metabolites were selected in the optimal Lasso model ($\beta < 0$), suggesting non-redundant associations with LTL. The coefficient of determination, a measure of goodness of fit, of the final model was estimated at 14.5% in a leave-one-out validation.

Moreover, we found all five metabolites to be strongly associated with several aging-related phenotypes independently of chronological age (Table 3): Both lysolipids correlated with increased systolic blood pressure (1-stearoylglycerophosphoinositol: Beta=1.09 [0.56:1.61], $P=5.3\times10^{-5}$ and 1-palmitoylglycerophosphoinositol: Beta=1.10 [0.52:1.67], $P=1.7\times10^{-4}$). Additionally, 1-palmitoyl-glycerophosphoinositol was found to be associated with the serum concentration of gamma-glutamyl transpeptidase (GGT), a measure of liver function (Beta=0.08 [0.03:0.12], $P=1.0\times10^{-3}$).

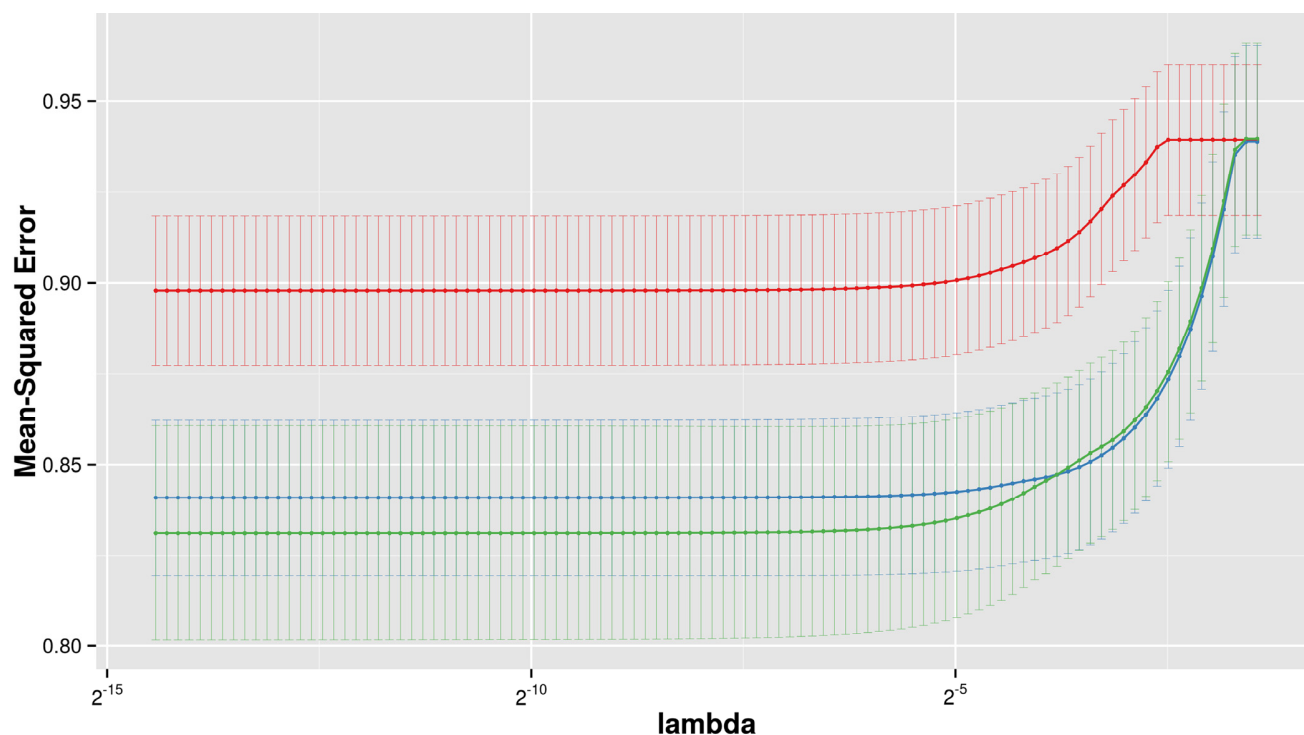


Figure 2. LTL prediction performance. The figure shows the prediction performance (mean square error on Y axis) of three different Lasso models, based on metabolites only (red), clinical variables only (blue) and metabolites with clinical variables combined (green), dependent on the amount of regularization (lambda on x axis).

Table 3. Phenotypes associated with LTL and associated metabolites.

	phenotype	beta [95%CI]	p
telomere length	HDL cholesterol	0.04 [0.02:0.06]	2.50×10^{-6}
	eGFR	1.42 [0.82:2.01]	2.79×10^{-6}
	smoking	-0.06 [-0.08:-0.03]	3.17×10^{-5}
	FEV1	0.03 [0.01:0.05]	8.85×10^{-4}
1-palmitoylglycerophosphoinositol	SBP	1.10 [0.52:1.67]	1.76×10^{-4}
	GGT	0.08 [0.03:0.12]	1.04×10^{-3}
1-stearoylglycerophosphoinositol	SBP	1.09 [0.56:1.61]	5.34×10^{-5}
4-vinylphenol sulfate	smoking	0.24 [0.22:0.26]	2.32×10^{-102}
	FEV1	-0.02 [-0.04:-0.01]	1.40×10^{-3}
gamma-glutamylphenylalanine	eGFR	-2.24 [-2.74:-1.73]	3.14×10^{-18}
	GGT	0.15 [0.10:0.19]	3.21×10^{-12}
	FEV1	-0.03 [-0.05:-0.02]	4.48×10^{-6}
	HDL cholesterol	-0.03 [-0.05:-0.02]	1.15×10^{-5}
	ALAT	0.10 [0.05:0.14]	5.76×10^{-5}
gamma-glutamyltyrosine	GGT	0.14 [0.10:0.19]	5.41×10^{-11}
	eGFR	-1.65 [-2.19:-1.11]	1.58×10^{-9}
	ALAT	0.11 [0.06:0.16]	1.67×10^{-5}

The two gamma-glutamyl amino acids were strongly associated with the estimated glomerular filtration rate (eGFR), a marker for renal function (gamma-glutamyltyrosine: Beta=-1.65 [-2.19:-1.11], $P=1.6 \times 10^{-9}$ and gamma-glutamylphenylalanine: Beta=-2.24 [-2.74:-1.73], $P=3.1 \times 10^{-18}$), and two markers of liver function, namely GGT and alanine amino transaminase (ALAT) (GGT: Beta=0.14 [0.10:0.19], $P=5.4 \times 10^{-11}$ and Beta=0.15 [0.10:0.19], $P=3.2 \times 10^{-12}$ respectively; ALAT: Beta=0.11 [0.06:0.16], $P=1.7 \times 10^{-5}$ and Beta=0.10 [0.05:0.14], $P=5.8 \times 10^{-5}$ respectively). Gamma-glutamylphenylalanine was additionally associated with lung function, measured as forced expiratory volume in one second (FEV1, Beta=-0.03 [-0.05:-0.02], $P=4.5 \times 10^{-6}$), and HDL cholesterol levels (Beta=-0.03 [-0.05:-0.02], $P=1.1 \times 10^{-5}$).

Moreover, the xenobiotic 4-vinylphenol sulfate was strongly associated with tobacco smoking (Beta=0.24 [0.22:0.26], $P=2.3 \times 10^{-102}$) and also weakly with FEV1 (Beta=-0.02 [-0.04:-0.01], $P=1.4 \times 10^{-3}$).

Thus, all five metabolites were consistently associated with accelerated biological aging, i.e. shorter telomeres, higher blood pressure and higher HDL cholesterol levels

and poorer lung, liver and kidney function (Table 3 and Figure 1).

To further investigate mechanisms of biological aging, we analyzed the association of the five significant metabolites with gene expression levels of related enzymes, namely GGT and phospholipase A2 (PLA2), in a subset of 753 individuals with RNA chip data from available LCL probes available. We found gamma-glutamyltyrosine was positively associated with GGT1 and GGTL3 gene expressions (probes ILMN_2274240: Beta=0.09 [0.02:0.15], $P=0.01$ and ILMN_1786186: Beta=0.07 [0.00:0.14], $P=0.04$). Also, 1-stearoylglycerophosphoinositol was positively associated with expression of the PLA2 gene PLA2G15 (probe ILMN_1756910: Beta=0.09 [0.01:0.16], $P=0.02$ and probe ILMN_1798955: Beta=0.08 [0.00:0.15], $P=0.05$) as well as 1-palmitoylglycerophosphoinositol (probe ILMN_1756910: Beta=0.08 [0.00:0.16], $P=0.05$).

The metabolite 4-vinylphenol sulfate is known to be associated with several DNA methylation probes, possibly driven by tobacco smoking [21,22]. We found one of these probes, cg19572487, being significantly

associated with LTL (Beta= 0.10 [0.04:0.17], $P=9\times 10^{-3}$), smoking (Beta=-0.51 [-0.63:-0.39], $P=9\times 10^{-16}$) and 4-vinylphenol sulfate levels (Beta=-0.05 [-0.09:-0.02], $P=1\times 10^{-3}$) in our data. The probe is located on chromosome 17 in the retinoic acid receptor, alpha (RARA) gene.

DISCUSSION

In the largest study of this kind, we searched for molecular markers and mechanisms involved in LTL regulation using a metabolomics approach. We identified five novel blood metabolites, namely gamma-glutamyltyrosine, gamma-glutamylphenylalanine, 1-stearoylglycerophosphoinositol, 1-palmitoylglycerophosphoinositol and 4-vinylphenol sulfate, independently associated with LTL with high statistical significance. These metabolites belong to three different classes: lysolipids, gamma-glutamyl amino acids and xenobiotics, which will be discussed in the following.

Lysolipids

Lysolipids are produced from glycerophospholipids by the enzyme phospholipase A2 (PLA2), which releases one of the fatty acids from the glycerol backbone [23]. Glycerophospholipids were previously found to be positively correlated with LTL [20] while in our study, circulating levels of the lysolipids 1-stearoylglycerophosphoinositol and 1-palmitoylglycerophosphoinositol were significantly associated with shortening of LTL. This suggests an increased activity of PLA2 in advanced biological aging. This hypothesis is further confirmed by the positive association of the two lysolipids with PLA2 gene expression levels in LCLs in our study. PLA2 activity, amongst others, affects the composition and physiology of cell membranes by catalyzing the hydrolysis of membrane lipids [24,25]. The integrity of cell membranes and their ability to resist oxidative stress have been shown to be key aspects of biological aging [26]. Studies comparing centenarians with younger controls identified alterations of cell membrane composition [27] and particularly depletion of the lysolipid stearylphosphatidylcholine [28] as possible reasons for longevity.

Another regulator of membrane fluidity is the saturation of fatty acids. Both stearic acid and palmitic acid are saturated fatty acids that are known to decrease membrane fluidity, which in turn was associated with increased susceptibility to disease [29,30,31]. In contrast higher levels of polyunsaturated fatty acid-containing phospholipids were observed in centenarians compared to elderly [32], suggesting their involvement in retarded biological aging. These alterations of

membrane composition with biological aging provide a possible explanation for previously reported association of LTL with e.g. AD [4].

Gamma-glutamyl amino acids

We found two gamma-glutamyl amino acids, gamma-glutamyltyrosine and gamma-glutamylphenylalanine, were negatively associated with LTL. These metabolites are components of the gamma-glutamyl cycle and are produced by the degradation of glutathione (GSH) and its conjugates catalyzed by the enzyme GGT. The main purpose of this reaction is regeneration of the intracellular GSH pool, i.e. to break-down extra-cellular GSH conjugates to make its components available for reimport into the cell [33,34,35]. GSH is crucial for detoxification of reactive oxygen species (ROS) as well as other toxic compounds [33,34,35]. Thus, increased GGT activity was proposed as a marker for increased oxidative stress [33,36]. Gamma-glutamyltyrosine and gamma-glutamylphenylalanine were both highly correlated with the abundance of the GGT enzyme, as well as GGT1 and GGTL3 gene expression in this study. The serum concentration of GGT is a common clinical marker for liver function [37]. While the liver produces most of the GSH [34], in the body GGT is most active in kidneys, which absorb GSH for detoxification [34,37]. Accordingly, we also found kidney function, measured as eGFR, being highly correlated with both, LTL and gamma-glutamyl amino acids. In conclusion, the gamma-glutamyl amino acids indicate an involvement of increased oxidative stress and worsened liver and kidney function in biological aging.

We also found gamma-glutamylphenylalanine being associated with worsened lung function in both cohorts. This might also be due to oxidative stress, which was previously associated with chronic lung disease [38].

4-Vinylphenol sulfate

4-vinylphenol sulfate is a xenobiotic that was reported to be strongly associated with tobacco smoking [39]. We observed the same correlation in our data. Moreover, we found both 4-vinylphenol sulfate as well as LTL to be strongly correlated with cotinine abundance, which is a well-established marker for tobacco smoking. Accordingly, higher levels of 4-vinylphenol sulfate were associated with worsened lung function. Moreover, analysis of DNA methylation data from our cohort confirmed previously published associations of 4-vinylphenol sulfate with the methylation level of a CpG site in the RARA gene [21] and revealed an association of the same site with LTL

and smoking. RARA is a transcription factor that was shown to regulate differentiation and apoptosis [40]. However, despite the strong correlation between LTL and smoking, we did not find a significant difference in LTL between monozygotic twins, discordant for smoking. These associations show how smoking accelerates biological aging mediated by changes in metabolism as well as DNA methylation. Smoking was shown to have a profound effect on the GSH metabolism of the lung [38], suggesting increased oxidative stress as a possible link between smoking, metabolism and LTL.

While we were able to identify five novel markers of LTL, our study has some limits. First, we analyzed data of females only and some of the identified metabolites are known to show gender-specific blood levels [41]. However, in a small pilot (n=372) from the TwinsUK cohort we observed concordant correlations between LTL and gamma-glutamyl amino acids as well as 4-vinylphenol sulfate for men as for women. In contrast, we did not see an association between any of the lysolipids and LTL in men, suggesting gender-specific changes of fatty acid metabolism with aging. Second, we did not reach statistical significance in the replication cohort. This can be attributed to smaller sample size. The power to detect the observed effects at a significance level of 0.05 in 900 individuals is only around 50%. Nonetheless, despite the lack of power, the much higher age and the different geographical location and genetic background of the replication cohort, all of the five metabolites remain Bonferroni-significant after meta-analysis.

Our results suggest two mechanisms of biological aging: On the one hand, changes in lipid metabolism and resulting changes of the cell membrane composition appear to be related to LTL and biological aging. On the other hand, we observed metabolites indicating increased oxidative stress due to alterations in the GSH metabolism, which has been previously related to LTL and aging phenotypes. One possible cause for increased oxidative stress is tobacco smoking, which might mediate the association of 4-vinylphenol sulfate with LTL. Moreover we found LTL and the related metabolites being associated with impairment of liver and kidney function. This highlights the importance of detoxification, particularly of reactive oxygen species, in biological aging.

METHODS

Discovery population. Study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins recruited as volunteers without selecting for

any particular disease or trait [42]. In this study we analyzed data from 3511 female twins who had who had complete data for LTL and metabolomics profiling. The study was approved by St. Thomas' Hospital Research Ethics Committee, and all twins provided informed written consent.

Replication cohort. KORA F4 is a population cohort based in the region of Augsburg, Germany [43]. The replication set consisted of 904 female individuals with serum metabolite levels, measures of telomeres and measures of lung function [44] available.

LTL measurement. A detailed description of LTL measurement in both TwinsUK and KORA was previously described in Codd et al. [15]. In brief, mean LTL of the samples was measured using a quantitative PCR-based technique [14,45] and expressed as a ratio of telomere repeat length (T) to a copy number of a single copy gene (S). A calibrator sample or a standard curve was used for to standardize T/S results across plates. LTLs measures were inverse normalized in both cohorts.

Metabolomics measurement. Metabolomics data was measured by Metabolon Inc., Durham, USA as previously described [46]. Briefly, metabolite concentrations were measured in blood samples using an untargeted GC/MS and LC/MS platform. Measurements were scaled by run-day median and inverse normalized in both cohorts.

Aging phenotypes. Lung function was measured as forced expiratory volume in one second (FEV1) in line with ATS/ERS recommendations [44,47]. Furthermore, diastolic and systolic blood pressure (DBP and SBP), body mass index (BMI) and serum HDL cholesterol levels were measured during clinical visits of the study participants. Renal function was measured by estimating glomerular filtration rate (eGFR) from serum creatinine levels using the CKD-EPI equation [48]. Liver function was assessed by measuring serum gamma-glutamyl transpeptidase (GGT) and alanine amino transaminase (ALAT) concentrations. Both measures were inverse normalized prior to analysis.

Gene expression. RNA abundance was measured in LCLs of 778 female individuals from the TwinsUK cohort using the Illumina Human HT-12 V3 BeadChip as part of the MuTHER project as previously described [49]. We selected 30 probes from GGT and PLA2 genes. Probes were adjusted for batch effects by linear models and residuals were inverse normalized prior to analysis.

DNA methylation. DNA was extracted from whole blood, bisulfite converted and subsequently analyzed

using the Infinium 450K kit as previously described [50]. The beta mixture quantile dilation (BMIQ) method was performed to correct for technical variation [51]. Measurements were inverse normalized and then adjusted for batch effects, family structure and cell counts (PlasmaBlast, CD8+CD28-CD45RA- T cells, naive CD8 T cells, CD4+ T cells, Natural Killer cells, monocytes, and granulocytes) using linear models.

Statistical analysis. All analyses were performed using R (version 3.1.2) using the lme4 (version 1.1) package.

Correlations between metabolites and LTL were calculated using linear mixed models, correcting for age, BMI and family relatedness (as random intercept). A conservative multiple test-corrected threshold of $P < 1.8 \times 10^{-4}$ was used to identify significant associations; this value represented $P = 0.05$ divided by the total number of tests performed (280 metabolites). We replicated the five Bonferroni-significant metabolites in the KORA F4 cohort. The data was consistently normalized in both cohorts. The results were meta-analyzed using inverse variance fixed effect meta-analysis implemented in the R package meta (version 4.3).

We estimated the power of the replication cohort using the R package pwr (version 1.1), which implements power estimation according to Cohen [52].

To identify redundant associations of the metabolites, we fitted a multivariate Lasso model [53] incorporating all Bonferroni significant metabolites together with age and BMI. The predictive performance of the model was then compared to a similar model containing age and BMI only. The model performance was assessed by calculating the predicted residual sum of squares (PRESS) and subsequent P^2 statistics using a leave-one-out cross validation.

Subsequently, we aimed to further explore the relationship of LTL and the identified metabolites with biological aging. To this end, we used linear mixed models to test for association of the previously identified metabolites with previously described aging phenotypes. All models were adjusted for age, BMI and family relatedness. The lung function parameter FEV1 was additionally adjusted for height, as suggested in the literature. We replicated the associations with lung function parameters in KORA, adjusting for the same covariates.

Funding

This work was supported by the EU Framework Programme 7 small-scale focused research collaborative

project EurHEALTHAging [277849]; TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme [FP7/2007-2013]. The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TDS is an ERC Advanced Researcher.

The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. This study was partly supported by the Competence Network Asthma and COPD (ASCONET), network COSYCONET (subproject 2, BMBF, FKZ 01GI0882) funded by the German Federal Ministry of Education and Research (BMBF). KS was supported by the Biomedical Research Program at Weill Cornell Medical College in Qatar, funded by the Qatar Foundation.

ACKNOWLEDGEMENTS

The authors would like to thank Alessia Visconti and Mario Falchi for their statistical advice.

Conflict of interest statement

Robert Mohny is an employee of Metabolon Inc.

REFERENCES

1. Lange T De. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes and Development*. 2005; 19:2100–2110.
2. Harley CB, Vaziri H, Counter CM, and Allsopp RC. The telomere hypothesis of cellular aging. *Experimental gerontology*. 1992; 27:375–382.
3. Valdes AM, Andrew T, Gardner JP, Kimura M, Oelsner E, Cherkas LF, Aviv A, and Spector TD. Obesity, cigarette smoking, and telomere length in women. *Lancet*. 2005; 366:662–664.
4. Panossian L. Telomere shortening in T cells correlates with Alzheimer's disease status. *Neurobiology of Aging*. 2003; 24:77–84.
5. Thomas P, O' Callaghan NJ, and Fenech M. Telomere length in white blood cells, buccal cells and brain tissue and its variation with ageing and Alzheimer's disease. *Mechanisms of Ageing and Development*. 2008; 129:183–190.
6. Fitzpatrick AL, Kronmal R a., Gardner JP, Psaty BM, Jenny NS, Tracy RP, Walston J, Kimura M, and Aviv A. Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *American Journal of Epidemiology*. 2007; 165:14–21.
7. Brouillette S, Singh RK, Thompson JR, Goodall AH, and Samani NJ. White cell telomere length and risk of premature myocardial infarction. *Arteriosclerosis, Thrombosis, and Vascular Biology*.

2003; 23:842–846.

8. Shay JW and Wright WE. Role of telomeres and telomerase in cancer. *Seminars in Cancer Biology*. 2011; 21:349–353.

9. Artandi SE and DePinho R a. Telomeres and telomerase in cancer. *Carcinogenesis*. 2009; 31:9–18.

10. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, and Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet*. 2003; 361:393–395.

11. Kimura M, Hjelmborg JVB, Gardner JP, Bathum L, Brimacombe M, Lu X, Christiansen L, Vaupel JW, Aviv A, and Christensen K. Telomere length and mortality: A study of leukocytes in elderly danish twins. *American Journal of Epidemiology*. 2008; 167:799–806.

12. Vera E, Bernardes de Jesus B, Foronda M, Flores JM, and Blasco M a. The rate of increase of short telomeres predicts longevity in mammals. *Cell reports*. 2012; 2:732–737.

13. Mather KA, Jorm AF, Parslow RA, and Christensen H. Is telomere length a biomarker of aging? A review. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. 2011; 66 A:202–213.

14. Codd V, Mangino M, Harst P van der, Braund PS, Kaiser M, Beveridge AJ, Rafelt S, Moore J, Nelson C, Soranzo N, Zhai G, Valdes AM, Blackburn H, et al. Common variants near TERC are associated with mean telomere length. *Nature genetics*. 2010; 42:197–199.

15. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I, Broer L, Nyholt DR, Leach IM, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics*. 2013; 45:422–427.

16. Mangino M, Christiansen L, Stone R, Hunt SC, Horvath K, Eisenberg DT a., Kimura M, Petersen I, Kark JD, Herbig U, Reiner a. P, Benetos A, Codd V, et al. DCAF4, a novel gene associated with leucocyte telomere length. *Journal of Medical Genetics*. 2015; 52:157–162.

17. Mangino M, Hwang S-J, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, Christiansen L, Petersen I, Elbers CC, Harris T, Chen W, Srinivasan SR, Kark JD, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Human Molecular Genetics*. 2012; 21:5385–5394.

18. Holmes E, Wilson ID, and Nicholson JK. Metabolic Phenotyping in Health and Disease. *Cell*. 2008; 134:714–717.

19. Hollywood K, Brison DR, and Goodacre R. Metabolomics: Current technologies and future trends. *Proteomics*. 2006; 6:4716–4723.

20. Zhao J, Zhu Y, Uppal K, Tran VT, Yu T, Lin J, Matsuguchi T, Blackburn E, Jones D, Lee ET, and Howard B V. Metabolic profiles of biological aging in american indians: The strong heart family study. *Aging (Albany, NY)*. 2014; 6:176–186.

21. Petersen a.-KAK, Zeilinger S, Kastenmüller G, Werner RM, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P, Huber F, Mohny RP, Grallert H, et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Human Molecular Genetics*. 2014; 23:534–545.

22. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, and Illig T. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS*

ONE. 2013; 8:e63812.

23. Dennis E. Diversity of group types, regulation, and function of phospholipase A2. *Journal of Biological Chemistry*. 1994; 269:13057–13060.

24. Farooqui, AA. and Horrocks, LA. *Glycerophospholipids in the Brain*. Springer New York: New York, NY 2007.

25. Cantin B, Brun LD, Gagné C, Murthy MR, Lupien PJ, and Julien P. Alterations in erythrocyte membrane lipid composition and fluidity in primary lipoprotein lipase deficiency. *Biochimica et biophysica acta*. 1992; 1139:25–31.

26. Hulbert A. Life, death and membrane bilayers. *The Journal of experimental biology*. 2003; 206:2303–2311.

27. Rabini R, Moretti N, Staffolani R, Salvolini E, Nanetti L, Franceschi C, and Mazzanti L. Reduced susceptibility to peroxidation of erythrocyte plasma membranes from centenarians. *Experimental Gerontology*. 2002; 37:657–663.

28. Collino S, Montoliu I, Martin FPJ, Scherer M, Mari D, Salvioli S, Bucci L, Ostan R, Monti D, Biagi E, Brigidi P, Franceschi C, and Rezzi S. Metabolic Signatures of Extreme Longevity in Northern Italian Centenarians Reveal a Complex Remodeling of Lipids, Amino Acids, and Gut Microbiota Metabolism. *PLoS ONE*. 2013; 8:1–12.

29. Kamada T and Otsuji S. Lower levels of erythrocyte membrane fluidity in diabetic patients. A spin label study. *Diabetes*. 1983; 32:585–591.

30. Aozaki S. Decreased membrane fluidity in erythrocytes from patients with Crohn's disease. *Gastroenterol.Jpn*. 1989; 24:246–254.

31. Mecocci P, Cherubini A, Beal MF, Cecchetti R, Chionne F, Polidori MC, Romano G, and Senin U. Altered mitochondrial membrane fluidity in AD brain. *Neuroscience Letters*. 1996; 207:129–132.

32. Caprari P, Scuteri A, Salvati AM, Bauco C, Cantafora A, Masella R, Modesti D, Tarzia A, and Marigliano V. Aging and red blood cell membrane: A study of centenarians. *Experimental Gerontology*. 1999; 34:47–57.

33. Zhang H, Forman HJ, and Choi J. Gamma-Glutamyl Transpeptidase in Glutathione Biosynthesis. *Methods in Enzymology*. 2005; 401:468–483.

34. Wu G, Fang Y-Z, Yang S, Lupton JR, and Turner ND. Glutathione metabolism and its implications for health. *The Journal of nutrition*. 2004; 134:489–492.

35. Maher P. The effects of stress and aging on glutathione metabolism. *Ageing Research Reviews*. 2005; 4:288–314.

36. Pandur S, Pankiv S, Johannessen M, Moens U, and Huseby N-E. Gamma-glutamyltransferase is upregulated after oxidative stress through the Ras signal transduction pathway in rat colon carcinoma cells. *Free radical research*. 2007; 41:1376–1384.

37. Whitfield JB. Gamma glutamyl transferase. *Critical reviews in clinical laboratory sciences*. 2001; 38:263–355.

38. Rahman I and MacNee W. Lung glutathione and oxidative stress: implications in cigarette smoke-induced airway disease. *The American journal of physiology*. 1999; 277:L1067–L1088.

39. Manini P, Palma G De, Andreoli R, Goldoni M, Poli D, Lasagni G, and Mutti A. Urinary excretion of 4-vinyl phenol after experimental and occupational exposure to styrene. *Giornale italiano di medicina del lavoro ed ergonomia*. 2003; 25 Suppl:61–62.

40. Hu X-X, Zhong L, Zhang X, Gao Y-M, and Liu B-Z. NLS-RAR α Promotes Proliferation and Inhibits Differentiation in HL-60 Cells. *International Journal of Medical Sciences*. 2014; 11:247–254.

41. Krumsiek J, Mittelstrass K, Do KT, Stücker F, Ried J, Adamski J, Peters A, Illig T, Kronenberg F, Friedrich N, Nauck M, Pietzner M, Mook-Kanamori DO, et al. Gender-specific pathway differences in the human serum metabolome. *Metabolomics*. 2015; 11:1815-1833.
42. Moayyeri A, Hammond CJ, Valdes AM, and Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. *International journal of epidemiology*. 2013; 42:76-85.
43. Holle R, Happich M, Löwel H, and Wichmann H. KORA - A Research Platform for Population Based Health Research. *Das Gesundheitswesen*. 2005; 67:19-25.
44. Karrasch S, Flexeder C, Behr J, Holle R, Huber RM, Jörres RA, Nowak D, Peters A, Wichmann H-E, Heinrich J, Schulz H, and KORA Study Group. Spirometric reference values for advanced age from a South german population. *Respiration; international review of thoracic diseases*. 2013; 85:210-219.
45. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic acids research*. 2002; 30:e47.
46. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *International journal of epidemiology*. 2013; 42:1111-1119.
47. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates a., Crapo R, Enright P, Grinten CPM van der, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, et al. Standardisation of spirometry. *European Respiratory Journal*. 2005; 26:319-338.
48. Levey AS, Stevens L a, Schmid CH, Zhang YL, Castro AF, Feldman HI, Kusek JW, Eggers P, Lente F Van, Greene T, and Coresh J. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*. 2009; 150:604-612.
49. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature genetics*. 2012; 44:1084-1089.
50. Tsai P-C, Dongen J Van, Tan Q, Willemsen G, Christiansen L, Boomsma DI, Spector TD, Valdes AM, and Bell JT. DNA methylation changes in the IGF1R gene in birth weight discordant adult monozygotic twins. *Twin Research and Human Genetics*. 2015; (in press)
51. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, and Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics (Oxford, England)*. 2013; 29:189-196.
52. Cohen, J. *Statistical power analysis for the behavioral sciences*. Routledge: Hillsdale, New Jersey 1988.
53. Tibshirani R. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*. 1994; 58:267-288.

G.3 Glycosylation Profile of IgG in Moderate Kidney Dysfunction

Glycosylation Profile of IgG in Moderate Kidney Dysfunction

Clara Barrios,^{*,†} Jonas Zierer,^{*,‡} Ivan Gudelj,[§] Jerko Štambuk,[§] Ivo Ugrina,[§] Eva Rodríguez,[†] María José Soler,[†] Tamara Pavić,^{||} Mirna Šimurina,^{||} Toma Keser,^{||} Maja Pučić-Baković,[§] Massimo Mangino,^{*} Julio Pascual,[†] Tim D Spector,^{*} Gordan Lauc,^{§||} and Cristina Menni^{*}

^{*}Department of Twin Research and Genetic Epidemiology, Kings College London, London, United Kingdom;

[†]Department of Nephrology, Hospital del Mar, Institut Mar d'Investigacions Mèdiques, Barcelona, Spain; [‡]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum Munich, Germany; [§]Genos Glycoscience Research Laboratory, Zagreb, Croatia; and ^{||}University of Zagreb, Faculty of Pharmacy and Biochemistry, Zagreb, Croatia

ABSTRACT

Glycans constitute the most abundant and diverse form of the post-translational modifications, and animal studies have suggested the involvement of IgG glycosylation in mechanisms of renal damage. Here, we explored the associations between IgG glycans and renal function in 3274 individuals from the TwinsUK registry. We analyzed the correlation between renal function measured as eGFR and 76 N-glycan traits using linear regressions adjusted for covariates and multiple testing in the larger population. We replicated our results in 31 monozygotic twin pairs discordant for renal function. Results from both analyses were then meta-analyzed. Fourteen glycan traits were associated with renal function in the discovery sample ($P < 6.5 \times 10^{-4}$) and remained significant after validation. Those glycan traits belong to three main glycosylation features: galactosylation, sialylation, and level of bisecting N-acetylglucosamine of the IgG glycans. These results show the role of IgG glycosylation in kidney function and provide novel insight into the pathophysiology of CKD and potential diagnostic and therapeutic targets.

J Am Soc Nephrol 27: ●●●–●●●, 2015. doi: 10.1681/ASN.2015010109

Chronic kidney disease affects 13% of the adult population in developed countries and it is associated with increased cardiovascular morbidity and mortality.^{1,2} Though many genetic^{3–5} and environmental factors (such as diabetes, hypertension and ageing)⁶ are implicated in the development of kidney damage, its pathophysiology is still not fully understood. Heritability estimates for CKD range between 0.33 and 0.41^{7,8} and despite the discovery of several important genetic associations, these loci collectively account for only 1.4% of the variation in eGFR.⁵ This suggests that epigenetic or post-transcriptional factors may be playing an important role in renal damage.

Glycosylation is the most abundant and diverse form of post-transcriptional modification and participates in every physiologic process.⁹

Immunoglobulin G is an excellent glycoprotein model as its glycosylation is well defined and many important functional effects of alternative IgG

glycosylation have been described.¹⁰ N-glycans attached to the conserved asparagine 297 in the Fc part of IgG are important modulators of IgG effector functions.¹¹ For example, glycosylation acts as a switch between pro- and anti-inflammatory IgG functionality. Malfunction of this system is associated with different inflammatory and autoimmune

Received January 30, 2015. Accepted June 5, 2015.

C.B., J.Z., G.L., C.M. and I.G. contributed equally to this work.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Clara Barrios, Department of Nephrology, Hospital del Mar, Institut Mar d'Investigacions Mèdiques, Barcelona, Spain, or Dr. Cristina Menni, Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK. E-mail: cristina.menni@kcl.ac.uk or CBarrios@parcdesalutmar.cat.

Copyright © 2015 by the American Society of Nephrology

diseases such as SLE,¹² rheumatoid arthritis, inflammatory bowel diseases,^{13,14} cancer^{15,16} and AIDS.¹⁷ Furthermore, it has been shown that inflammation pathways play a key role in endothelial and kidney damage.^{18,19} Indeed, the activation of inflammatory pathways and subsequent fibrosis are hallmark of renal injury.^{20,21} Different IgG glycosylation profiles may provide an at-risk phenotype to the development of renal damage.

Animal models highlighted the potential role of IgG glycosylation in the pathophysiologic mechanism involved in renal damage. Indeed studies have shown that modulation of ANCA IgG glycosylation reduces its pathogenicity in mouse ANCA-associated GN.²¹ Also, IgG Fcγ receptor deficiency was found to be renoprotective in a mouse model of diabetic nephropathy.²⁰ Human studies suggest that aberrant glycosylation of the IgA1 is implicated in the deposit and formation of the immunocomplex IgA–IgG in patients with IgA nephropathy.^{22,23}

However, no human studies investigated the role of the IgG glycosylation profiles in the onset of CKD.

The aim of this study is to investigate the potential role of IgG glycosylation in kidney function, by analyzing IgG glycome composition in a large population-based cohort from the UK. As glycans are associated with many factors including genes,²⁴ we validate our significant results in an independent population of identical twins discordant for renal diseases.

RESULTS

Levels of 76 IgG glycans (24 directly measured and 52 derived traits) (Supplemental Figure 1) were obtained in 3274 individuals with different eGFR from the TwinsUK population (age range: 18–87 years). The demographic characteristics of the study populations are presented in Table 1. We identified 31 monozygotic (MZ) twin pairs discordant for the renal phenotype (difference in eGFR > 15 mL/min per 1.73 m²).

We first ran the linear regressions in the discovery population adjusting for age, sex, body mass index (BMI), diabetes,

hypertension, glycan analysis batch and family relatedness, excluding the MZ discordant twins. We controlled for multiple testing using Bonferroni correction ($P < 6.5 \times 10^{-4}$; 0.05/76 glycan traits). This identified 14 glycans significantly associated with eGFR; six glycans were positively associated with eGFR, while eight were negatively associated (Table 2, Supplemental Table 1). To ensure that sexual hormones did not affect our results, we ran the same linear regression analysis including menopause as a covariate and our results were unchanged.

We then assessed whether these associations with renal function were robust by testing an independent group of MZ twins discordant for renal disease. The regression coefficients were in the same direction in both analyses (discordant identical twins and the rest of the population). We then combined the results using inverse-variance fixed effect meta-analysis. All 14 glycans remained Bonferroni significant (Table 2). As depicted in Figure 1 and Table 2, the 14 significant glycan traits fell into three particular glycosylation features: galactosylation, sialylation and the level of bisecting *N*-acetylglucosamine (GlcNAc) of the IgG glycans.

We observed a decrease in agalactosylated glycans: A2 (GP2 and GP2ⁿ) and FA2B (GP6 and GP6ⁿ) glycan structures and derived trait G0ⁿ, which combines all agalactosylated structures. Conversely, glycan with galactose on both antennae, FA2G2 (GP14 and GP14ⁿ), and the G2ⁿ derived trait, representing the percentage of digalactosylated structures in neutral IgG glycans, increased in parallel with the eGFR. The same pattern was observed in the MZ discordant pairs. As for sialylation, the major sialylated glycan, FA2G2S1 (GP18) and the percentage of sialylated structures without bisecting GlcNAc (represented by the ratio FGS/[F+FG+FGS]) increased with eGFR.

The level of bisecting GlcNAc in sialylated IgG glycans represented by three ratios, FBS^{total}/FS^{total}, FBS1/FS1, and FBS1/(FS1+FBS1), as well as in digalactosylated neutral gG glycans (FG2ⁿ/[BG2ⁿ+FBG2ⁿ]) were found to be inversely associated with eGFR.

To reinforce our findings we searched for associations in an independent population with more severe renal phenotype (eGFR < 30 mL/min per 1.73 m²). Eight twins, mean aged 65.0 (range 42.2–75.5 years) with CKD stage 4/5 (mean eGFR 24.7 [range 8.0–27.3]) were compared with their age-matched co-twin with eGFR > 30 mL/min per 1.73 m². As depicted in Figure 2, IgG glycans profiles follow the same patterns as were observed in the discovery population with the worsening of the renal function.

To determine whether the findings were restricted to IgG or to a more general change in glycosylation of multiple proteins, we searched for association between total plasma glycome^{25,26} and eGFR in a subset 426 individuals (eGFR, mL/min/1.73 m²: 78.95 ± 16.00). We found no difference in plasma glycosylation, suggesting that the effects we see here are likely direct effects of IgG glycosylation. However, the lack of association might also be due to power issues and so further study on larger sample size is needed to test this (Supplementary Table 2).

Table 1. General characteristics of the study population

	Discovery Population	MZ Discordant Twins
Sample size, <i>n</i>	3212	62
Age, years	52.67 ± 14.15	55.45 ± 12.2
MZ:DZ:singleton	506:1772:934	62:0:0
Female, <i>n</i> (%)	3050 (94.9)	60 (96.7)
BMI, kg/m ²	25.95 ± 4.65	25.64 ± 5.65
Creatinine, mg/ml	0.83 ± 0.15	0.75 ± 0.10
eGFR, mL/min per 1.73 m ²	84.15 ± 17.02	88.52 ± 9.91
CKD (eGFR ≤ 60), <i>n</i> (%)	294 (9.15)	1 (1.6)
Type II diabetes, <i>n</i> (%)	72 (2.2)	4 (6.4)
Hypertension, <i>n</i> (%)	705 (21.9)	18 (29.0)

CKD eGFR estimated using Chronic Kidney Disease Epidemiology Collaboration equation. Values for categorical variables are given as *n* (%); values for continuous variable as mean (±SD). MZ:DZ, monozygotic:dizygotic.

Table 2. Glycan traits significantly associated with eGFR in the discovery, validation, and meta-analysis

Glycan	Description	h^2a	Discovery		P value	MZ Discordant		Fixed effect meta-analysis	
			β (95% CI)			β (95% CI)		β (95% CI)	P value
GP18	The percentage of FA2G2S1 glycan in total IgG glycans	0.73	1.48 (0.89 to 2.07)		8.60×10^{-7}	0.59 (−2.23 to 3.41)		4.23 (2.38 to 7.52)	9.51×10^{-7}
GP14	The percentage of FA2G2 glycan in total IgG glycans	0.36	1.46 (0.85 to 2.07)		2.92×10^{-6}	1.33 (−1.81 to 4.48)		4.29 (2.35 to 7.81)	2.04×10^{-6}
GP6 ⁿ	The percentage of FA2B glycan in total neutral IgG glycans (GP ⁿ)	0.75	−1.39 (−1.98 to −0.80)		3.56×10^{-6}	−0.84 (−3.44 to 1.76)		0.26 (0.14 to 0.45)	3.16×10^{-6}
GP14 ⁿ	The percentage of FA2G2 glycan in total neutral IgG glycans (GP ⁿ)	0.47	1.29 (0.68 to 1.90)		3.06×10^{-5}	1.99 (−1.70 to 5.67)		3.70 (2.03 to 6.73)	1.82×10^{-5}
FBS1/FS1	Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc	0.39	−1.12 (−1.65 to −0.59)		3.48×10^{-5}	−0.58 (−3.16 to 1.99)		0.33 (0.20 to 0.56)	3.42×10^{-5}
FBS1/(FS1+FBS1)	The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans	0.42	−1.10 (−1.63 to −0.57)		4.63×10^{-5}	−0.60 (−3.14 to 1.95)		0.34 (0.20 to 0.57)	4.46×10^{-5}
G2 ⁿ	The percentage of digalactosylated structures in total neutral IgG glycans	0.41	1.20 (0.60 to 1.80)		8.81×10^{-5}	1.98 (−1.83 to 5.78)		3.38 (1.87 to 6.10)	5.53×10^{-5}
GP6	The percentage of FA2B glycan in total IgG glycans	0.75	−1.14 (−1.71 to −0.57)		8.90×10^{-5}	−1.01 (−3.78 to 1.76)		0.32 (0.18 to 0.56)	6.84×10^{-5}
FBS ^{total} /FS ^{total}	Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc	0.23	−1.07 (−1.60 to −0.54)		8.21×10^{-5}	−0.30 (−2.84 to 2.23)		0.36 (0.21 to 0.60)	9.52×10^{-5}
G0 ⁿ	The percentage of agalactosylated structures in total neutral IgG glycans	0.72	−1.16 (−1.76 to −0.56)		1.52×10^{-4}	−1.20 (−4.71 to 2.31)		0.31 (0.17 to 0.57)	1.20×10^{-4}
GP2 ⁿ	The percentage of A2 glycan in total neutral IgG glycans (GP ⁿ)	0.71	−0.91 (−1.42 to −0.40)		5.02×10^{-4}	−2.00 (−4.66 to 0.67)		0.39 (0.23 to 0.64)	2.20×10^{-4}
GP2	The percentage of A2 glycan in total IgG glycans	0.72	−0.90 (−1.42 to −0.38)		6.28×10^{-4}	−2.33 (−5.13 to 0.47)		0.39 (0.23 to 0.64)	2.55×10^{-4}
FGS/(F+FG+FGS)	The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans	0.69	1.01 (0.46 to 1.56)		2.96×10^{-4}	0.57 (−2.21 to 3.35)		2.71 (1.58 to 4.64)	2.85×10^{-4}
FG2 ⁿ /(BG2 ⁿ + FBG2 ⁿ)	Ratio of fucosylated digalactosylated nonbisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc	0.66	0.91 (0.38 to 1.44)		7.32×10^{-4}	0.93 (−1.59 to 3.44)		2.49 (1.48 to 4.19)	5.54×10^{-4}

^aEstimates of heritability (h^2) come from Menni et al. *Plos One* 2013.²⁴

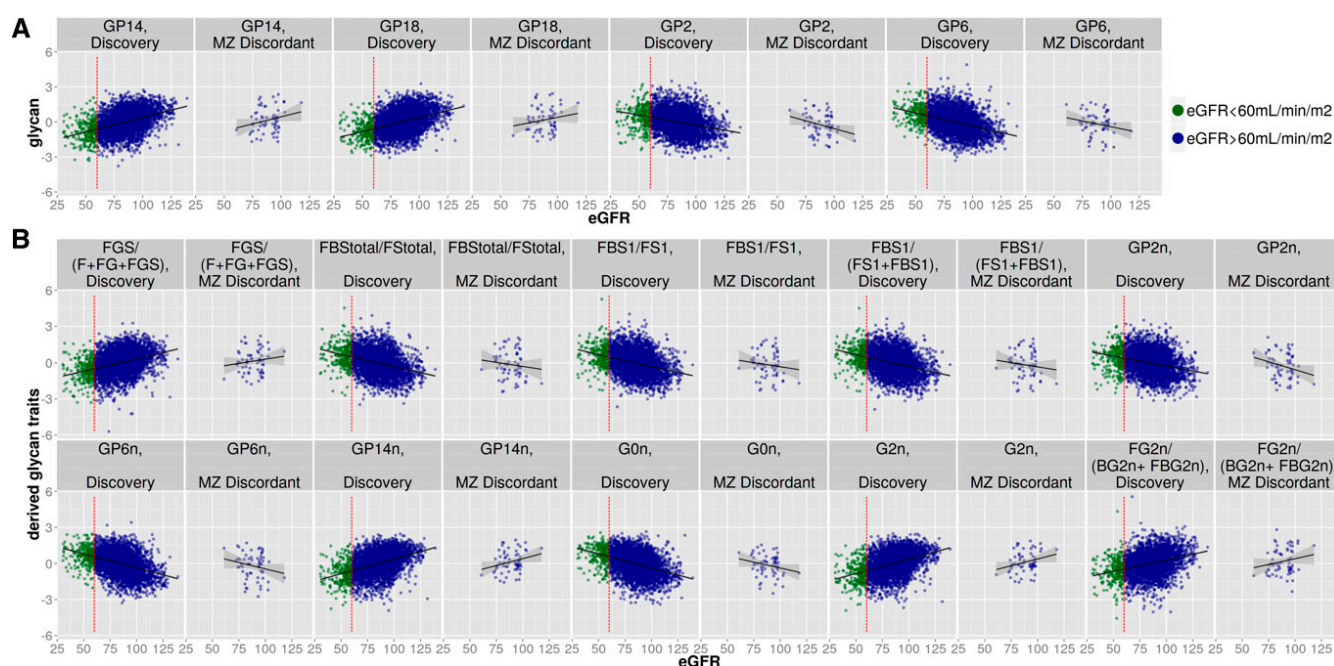


Figure 1. Correlation of IgG glycosylation and eGFR in the discovery and MZ discordant populations. (A) Directly measured glycan structures. (B) Derived traits that measure sialylation, galactosylation, and bisecting GlcNAc.

Finally, we assessed whether glycan profiles could improve the prediction of the CKD status (as per Guidelines, CKD cases have $\text{eGFR} < 60 \text{ mL/min per } 1.73 \text{ m}^2$) beyond that achieved with age and sex. In the discriminative model only the four main glycans (GP2, GP6, GP14, and GP18) were included. The predictive ability for CKD status, as measured by the area under the curve was 0.87 (95% confidence interval [95% CI], 0.85 to 0.89) for clinical parameters alone, 0.81 (95% CI, 0.78 to 0.84) for glycans alone, and 0.88 (95% CI, 0.86 to 0.90) for the model incorporating a combination of glycans and clinical parameters ($P=0.23$) (Supplemental Figure 2).

DISCUSSION

This is the first study to investigate the potential role of IgG glycosylation in kidney function. We identified 14 IgG glycan traits with high statistical significance associated with eGFR and validated them in an independent subset of MZ twins discordant for renal disease. Moreover we see the same pattern in a small independent sample with a more extreme renal dysfunction.

The glycans identified fall into three principal glycan traits.

Galactosylation of IgG

Decreased IgG galactosylation has been found to be associated with rheumatoid arthritis²⁷ as well as with several autoimmune and inflammatory diseases¹⁶ and with chronologic and biologic age.²⁸ The decrease in galactosylation is not disease-specific, but a general phenomenon that is associated with

decreased immunosuppressive and anti-inflammatory potential of circulating IgG. We observed a higher risk of CKD in subjects with agalactosylated glycans (GP2, GP6, and G0ⁿ) and lower in those with galactosylated IgG (GP14 and G2ⁿ). Lack of terminal galactose activates complement cascade and makes IgG pro-inflammatory, whereas the addition of galactose decreases its inflammatory potential.^{29,30} Hence, the IgG galactosylation pattern observed in our population supports the theory that complement activation/dysregulation is crucial in renal damage.³¹ It is not clear whether IgG galactosylation is a consequence or an individual predisposition for a disease. The heritability of galactosylated glycans was very high,²⁴ indicating that galactosylation could partly be genetically predetermined. This hypothesis is further supported by the fact that in rheumatoid arthritis, the decrease in IgG galactosylation was observed up to several years before the onset of the disease.^{32–35}

Sialylation

Further extension of IgG glycans by the addition of sialic acid dramatically changes the physiologic role of IgG, converting it from a proinflammatory into an anti-inflammatory agent.^{36,37} This relatively small fraction of sialylated IgG is believed to be responsible for the immunosuppressive activity of intravenously administered immunoglobulins.³⁸ Approximately 50% of IgG glycans are not sialylated and are proinflammatory.³⁹ However, the terminal $\alpha 2,6$ -sialylation of IgG glycans decreases the ability of IgG to bind Fc γ receptors (Fc γ Rs), which increases expression of inhibitory Fc γ RIIB and is anti-inflammatory.⁴⁰ Contrary to changes in galactosylation, the significant changes in sialylation have not been associated

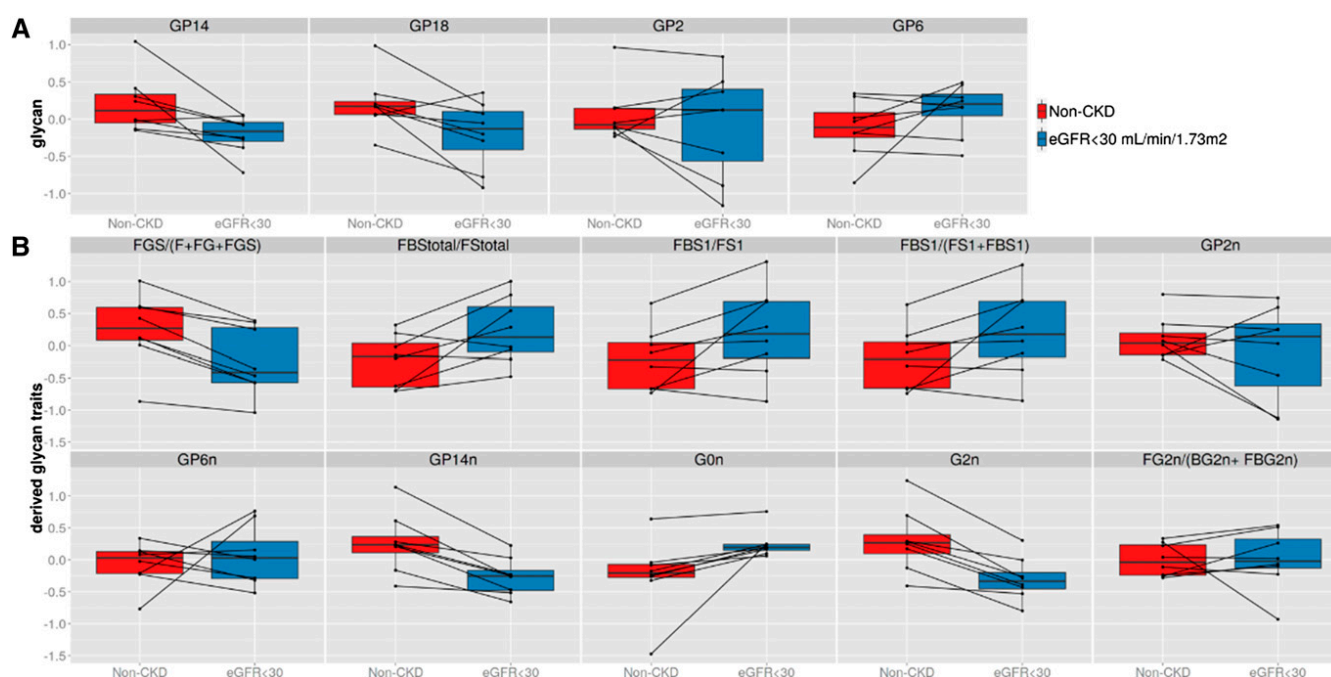


Figure 2. IgG glycan profiles in eight pairs of twins discordant for renal function. Comparisons between each pair of twins where one has extreme renal phenotype (eGFR < 30 mL/min per 1.73 m²) versus non-CKD. (A) Directly measured glycans structures. (B) Derived traits that measure sialylation, galactosylation, and bisecting GlcNAc. Results are in line with those observed in the discovery population (Figure 1).

with other diseases. Recently, some of us found that major sialylated glycans (GP16, GP18, and GP23) were significantly decreased in patients with SLE (F. Vučković *et al.*, submitted for publication). In our population, the major sialylated glycan, FA2G2S1 (GP18), and the ratio FGS/(F+FG+FGS), which represents the percentage of sialylated structures without bisecting GlcNAc in total IgG glycans, were decreased in patients with CKD (green dots in Figure 1). These sialylated glycan traits displayed a protective independent risk for CKD.

Bisecting *N*-Acetylglucosamine and Core Fucosylation of IgG

Another feature is the role of core fucose in the modulation of antibody-dependent cellular cytotoxicity.⁴¹ On average, 95% of the IgG population is core fucosylated⁴²; hence, most of the immunoglobulins have a “safety switch”, which prevents them from antibody-dependent cellular cytotoxicity. IgG-containing glycans that lack core fucose have 100-fold higher affinity to the FcγRIIIa and are therefore much more efficient than fucosylated glycoforms.⁴³ We have observed a significant and independent decreased risk of CKD when sialylated and core fucosylated glycans did not have bisecting GlcNAc; and in contrast, lower eGFR if those glycans contained bisecting GlcNAc (FBS^{total}/FS^{total}, FBS1/FS1, and FBS1/(FS1+FBS1)). Also for neutral digalactosylated glycans, when there is less of these glycans with bisecting GlcNAc, the ratio FG2ⁿ/(BG2ⁿ+FBG2ⁿ) is higher and this is positively associated with eGFR. The presence of bisecting GlcNAc was always associated with a higher risk of CKD.

It is not clear how the modulation of antibody-dependent cellular cytotoxicity could affect the renal damage in the onset of a nonautoimmune CKD. Studies in experimental animals have reported that modifications in the Fcγ receptor can diminish renal damage in a well known autoimmune disease, ANCA-related GN, as well as in diabetic nephropathy.^{20,21} On the other hand, renal fibrosis is the common pathway of many kidney diseases and leads to progressive renal failure; natural killer cells have been linked with this process in different organ systems.¹¹

Notably, glycan traits associated with lower eGFR have on average a higher heritability (Table 2). For example, the agalactosylated IgG glycans we found associated with lower eGFR, have a high heritability, ranging from 0.72 to 0.75, whereas galactosylated glycans GP14 and G2ⁿ derived trait have a low heritability (0.36 and 0.41, respectively).²⁴ The highly heritable glycans associated with eGFR, have been previously associated with different genes.¹² However, there is as yet no overlap with genes previously reported in CKD genome-wide association studies.⁵ Our findings may indicate a new approach to deeper understanding of the contribution of genetics in IgG glycosylation and kidney damage.

Although the identified glycans do not predict incident CKD (defined as eGFR < 60 mL/min per 1.73 m²) more accurately than clinical parameter, their inclusion in the models improves the incident CKD risk prediction. These glycans may be more sensitive to earlier stages of reduced renal function, as the eGFR-defined onset of CKD occurs only after half of the

kidneys' filtration ability has been lost. Longitudinal studies could help to address this hypothesis.

The present study has several strengths. First, we employed a two-stage design (discovery and independent replication with stringent *P* values), so minimizing the risk of false positive findings. Second, we used identical twins discordant for renal function in the validation analysis. Glycan levels may be influenced by many factors including genetics, age and environment.¹² As identical twins share 100% of their genetic makeup, and are matched perfectly for age, gender, social class, *etc.*, we were able to validate the role of IgG on renal function; isolating the nongenetic contribution. These data help us to understand the complex interplay between genetic and nongenetic influences that determine renal function.

We note some study limitations. First, there is a female predominance in our study sample (95% of the individuals are, for historical reasons, women). Second, our population being volunteers is slightly healthier than average with a lower rate of diabetes and results might not be generalizable to more severe diabetes populations. Third, the cross-sectional nature of our data does not allow us to draw conclusions as to whether the glycans identified are causative of kidney function decline or merely correlated with it. Finally, we cannot provide reliable estimates as to what proportions of the identified glycans were from Fc and from Fab, respectively. However, in a small pilot of Fc-glycopeptides by nano-liquid chromatography tandem mass spectrometry³⁹ on 96 representative age-matched individuals from the extremes of the eGFR distribution, we find the same direction of effect with renal function for all but one. This suggests that our initial observations mostly come from the Fc glycans (Supplemental Table 3).

Our results highlight the promising role of glycomics in renal studies. Uncovering this relationship by extending the research with clinical subsets and longitudinal data would help to identify further novel markers that would be potentially useful to detect at-risk patients, in the early stages of CKD. These results open new avenues to our understanding of renal damage and encourage further studies in populations with more severe CKD and proteinuria information, as well as studies comparing patients with autoimmune CKD with patients whose CKD is due to other etiologies. Moreover, this would help to gain additional insights into the pathophysiology of CKD and potential therapeutic targets.

CONCISE METHODS

Study Subjects

Study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins. Twins were recruited as volunteers by successive media campaigns without selecting for particular diseases or traits.⁴⁴ In this study we analyzed data from 3274 individuals with glycomics and creatinine data available. The study was approved by St. Thomas' Hospital Research Ethics Committee, and all twins provided informed written consent.

Phenotype Definitions

Data relevant to the present study include BMI (body weight in kilograms divided by the square of height in square meters), type II diabetes (defined if fasting glucose ≥ 7 mmol/L or physician's letter confirming diagnosis) and hypertension. Renal parameters; eGFR was calculated from standard creatinine using the Chronic Kidney Disease Epidemiology Collaboration equation.⁴⁵ CKD was defined as an eGFR < 60 ml/min per 1.73 m^2 according to the current Kidney Disease Outcome Quality Initiative (K/DOQI) guidelines.⁴⁶ MZ pairs were considered discordant for renal function if one twin had an eGFR ≥ 90 and the other had eGFR ≤ 90 mL/min per 1.73 m^2 and the difference between their eGFR levels was > 15 ml/min per 1.73 m^2 .

Analysis of IgG Glycans

Isolation of IgG from Human Plasma

The IgG was isolated using protein G monolithic plates (BIA Separations, Ajdovščina, Slovenia) as described previously.⁴²

Glycan Release and Labeling

Glycan release and labeling were performed essentially as previously described.^{24,42} Briefly, dried IgG was denatured with 2% SDS (wt/vol) and N-glycans were released by digestion with PNGase F (ProZyme, Hayward, CA). After deglycosylation, N-glycans were labeled with 2-AB fluorescent dye. Free label and reducing agent were removed from the samples using hydrophilic interaction chromatography–solid-phase extraction.

Hydrophilic Interaction Chromatography-UPLC

Fluorescently labeled N-glycans were separated by hydrophilic interaction chromatography on a Waters Acquity UPLC instrument (Waters, Milford, MA) as described previously.⁴² Data processing was performed using an automatic processing method with a traditional integration algorithm after which each chromatogram was manually corrected to maintain the same intervals of integration for all the samples. The chromatograms were all separated in the same manner into 24 peaks and the amount of glycans in each peak was expressed as a percentage of the total integrated area. In addition to 24 directly measured glycan structures, 52 derived traits were calculated. These derived traits average particular glycosylation features (galactosylation, fucosylation, bisecting GlcNAc, and sialylation) (Supplemental Figure 1, Table 1).

Statistical Analysis

Statistical analysis was carried out using Stata version 12 and R (version 3.1.2) and visualized using the ggplot2 package.

Glycans were globally normalized and log transformed using the right-skewness of their distributions. To remove experimental biases, all measurements were adjusted for batch and run-day effects using ComBat (R-package sva). Derived glycan traits were calculated using normalized and batch-corrected glycan measurements (exponential of batch corrected measurements). All variables were centered and scaled to have mean 0 and standard deviation 1. Outliers (more than 6SD from the mean) were excluded from the analysis.

Association analyses between eGFR and glycan traits were performed using random intercept linear regressions adjusting for age, sex, BMI, diabetes, hypertension, and family relatedness as

random effect. We used a conservative Bonferroni correction to account for multiple testing assuming 76 independent tests as suggested by Pucic *et al.*,⁴² so giving a significant threshold of ($P < 6.5 \times 10^{-4}$; 0.05/76). The Bonferroni-significant eGFR glycan associations were replicated in the previously excluded group of MZ discordant twins using the same model. Paired *t*-tests were used to evaluate the association with incident CKD in an independent subset of twins where one co-twin had a significant decline in renal function.

To assess how glycans can improve the prediction of CKD (eGFR < 60 ml/min per 1.73 m²), three Least Absolute Shrinkage and Selection Operator regression models were created (R package glmnet): The first one using only clinical parameters; age, sex, type II diabetes, and hypertension, to predict CKD, the second using the set of original glycan traits, which were found to be Bonferroni significant before (GP2, GP6, GP14, GP18), and the last one using both glycans and clinical parameters. The quality of all three models was assessed using a ten-fold cross-validation. The regularization parameter λ was trained separately for each fold using a nested cross-validation. Receiver operating characteristic curves (and particularly the area under the curves) were calculated for each fold and averages and confidence intervals were reported.

ACKNOWLEDGMENTS

TwinsUK was funded by the Wellcome Trust, European Community's Seventh Framework Programme (FP7/2007-2013).

The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St. Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust and King's College London.

Tim Spector is holder of an ERC Advanced Principal Investigator award.

Clara Barrios is supported by a grant from the Spanish Society of Nephrology and RedinRen RD12/0021/0024.

Julio Pascual, María José Soler, and Eva Rodríguez are supported by grants from Institute Carlos III and RedinRen.

Glycan analysis was partly funded by European Commission FP7 projects IBD-BIOM (contract #305479), HighGlycan (contract #278535), MIMOmics (contract #305280), HTP-GlycoMet (contract #324400), and IntegraLife (contract #315997) grants.

DISCLOSURES

Gordan Lauc is founder and owner of Genos, a private research organization that specializes in high-throughput glycomic analysis and has several patents in this field.

REFERENCES

- Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY: Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 351: 1296–1305, 2004
- Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, Van Lente F, Levey AS: Prevalence of chronic kidney disease in the United States. *JAMA* 298: 2038–2047, 2007
- Pattaro C, Köttgen A, Teumer A, Garnaas M, Böger CA, Fuchsberger C, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gieman HJ, Feitosa M, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson Å, Tönjes A, Dehghan A, Chouraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G, Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu FB, Demirkan A, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Koenig W, Illig T, Döring A, Wichmann HE, Kolcic I, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Endlich K, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Giulianini F, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, Thun GA, Paulweber B, Haun M, Sala C, Metzger M, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllenstein U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman JC, Hayward C, Ridker P, Parsa A, Bochud M, Heid IM, Goessling W, Chasman DI, Kao WH, Fox CS; CARDIoGRAM Consortium; ICBP Consortium; CARE Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2): Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* 8: e1002584, 2012
- Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, Takahashi A, Maeda S, Tsunoda T, Chen P, Lim SC, Wong TY, Liu J, Young TL, Aung T, Seielstad M, Teo YY, Kim YJ, Lee JY, Han BG, Kang D, Chen CH, Tsai FJ, Chang LC, Fann SJ, Mei H, Rao DC, Hixson JE, Chen S, Katsuya T, Isono M, Ogihara T, Chambers JC, Zhang W, Kooner JS, Albrecht E, Yamamoto K, Kubo M, Nakamura Y, Kamatani N, Kato N, He J, Chen YT, Cho YS, Tai ES, Tanaka T; KidneyGen Consortium; CKDGen Consortium; GUGC consortium: Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* 44: 904–909, 2012
- Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O'Connell JR, Li M, Schmidt H, Tanaka T, Isaacs A, Ketkar S, Hwang SJ, Johnson AD, Dehghan A, Teumer A, Paré G, Atkinson EJ, Zeller T, Lohman K, Cornelis MC, Probst-Hensch NM, Kronenberg F, Tönjes A, Hayward C, Aspelund T, Eiriksdottir G, Launer LJ, Harris TB, Rumpfer E, Mitchell BD, Arking DE, Boerwinkle E, Struchalin M, Cavalieri M, Singleton A, Giallauria F, Metter J, de Boer IH, Haritunians T, Lumley T, Siscovick D, Psaty BM, Zillikens MC, Oostra BA, Feitosa M, Province M, de Andrade M, Turner ST, Schillert A, Ziegler A, Wild PS, Schnabel RB, Wilde S, Munzel TF, Leak TS, Illig T, Klopp N, Meisinger C, Wichmann HE, Koenig W, Zgaga L, Zemunik T, Kolcic I, Minelli C, Hu FB, Johansson A, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Schreiber S, Aulchenko YS, Felix JF, Rivadeneira F, Uitterlinden AG, Hofman A, Imboden M, Nitsch D, Brandstätter A, Kollerits B, Kedenko L, Mägi R, Stumvoll M, Kovacs P, Boban M, Campbell S, Endlich K, Völzke H, Kroemer HK, Nauck M, Völker U, Polasek O, Vitart V, Badola S, Parker AN, Ridker PM, Kardia SL, Blankenberg S, Liu Y, Curhan GC, Franke A, Roach T, Paulweber B, Prokopenko I, Wang W, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Shlipak MG, van Duijn CM, Borecki I, Krämer BK, Rudan I, Gyllenstein U, Wilson JF, Witteman JC, Pramstaller

- PP, Rettig R, Hastie N, Chasman DI, Kao WH, Heid IM, Fox CS: New loci associated with kidney function and chronic kidney disease. *Nat Genet* 42: 376–384, 2010
6. Meguid El Nahas A, Bello AK: Chronic kidney disease: the global challenge. *Lancet* 365: 331–340, 2005
 7. Bochud M, Elston RC, Maillard M, Bovet P, Schild L, Shamlaye C, Burnier M: Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean). *Kidney Int* 67: 61–69, 2005
 8. Langefeld CD, Beck SR, Bowden DW, Rich SS, Wagenknecht LE, Freedman BI: Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am J Kidney Dis* 43: 796–800, 2004
 9. National Research Council: Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. Transforming Glycoscience: A Roadmap for the Future, Washington, D.C., The National Academies Press, 2012
 10. Gornik O, Pavić T, Lauc G: Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim Biophys Acta* 1820: 1318–1326, 2012
 11. Jang HR, Rabb H: Immune cells in experimental acute kidney injury. *Nat Rev Nephrol* 11: 88–101, 2015
 12. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, Novokmet M, Polašek O, Gornik O, Krištić J, Keser T, Vitart V, Scheijen B, Uh HW, Molokhia M, Patrick AL, McKeigue P, Kolčić I, Lukić IK, Swann O, van Leeuwen FN, Ruhaak LR, Houwing-Duistermaat JJ, Slagboom PE, Beekman M, de Craen AJ, Deelder AM, Zeng Q, Wang W, Hastie ND, Gyllenstein U, Wilson JF, Wuhler M, Wright AF, Rudd PM, Hayward C, Aulchenko Y, Campbell H, Rudan I: Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* 9: e1003225, 2013
 13. Go MF, Schrohenloher RE, Tomana M: Deficient galactosylation of serum IgG in inflammatory bowel disease: correlation with disease activity. *J Clin Gastroenterol* 18: 86–87, 1994
 14. Trbojević-Akmačić I, Ventham NT, Theodoratou E, Vuckovic F, Kennedy NA, Kristic J, Nimmo ER, Drummond D, Stambuk J, Klaric L, Dunlop MG, Novokmet M, Aulchenko Y, Gornik O, Kolarich D, Wuhler M, McGovern D, Annesse V, Kalla R, Pemberton JM, Spencer D, Zoldos V, Fernandes D, Campbell H, Pucic Bakovic M, Satsangi J, Lauc G: Inflammatory bowel disease associates with pro-inflammatory potential of the IgG glycome. *Inflamm Bowel Dis* 21(6): 1237–1247, 2015
 15. Ohtsubo K, Marth JD: Glycosylation in cellular mechanisms of health and disease. *Cell* 126: 855–867, 2006
 16. Gornik O, Lauc G: Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* 25: 267–278, 2008
 17. Moore JS, Wu X, Kulhavy R, Tomana M, Novak J, Moldoveanu Z, Brown R, Goepfert PA, Mestecky J: Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* 19: 381–389, 2005
 18. Paragh G, Seres I, Harangi M, Fülöp P: Dynamic interplay between metabolic syndrome and immunity. *Adv Exp Med Biol* 824: 171–190, 2014
 19. Camps J, García-Heredia A: Introduction: oxidation and inflammation, a molecular link between non-communicable diseases. *Adv Exp Med Biol* 824: 1–4, 2014
 20. Lopez-Parra V, Mallavia B, Lopez-Franco O, Ortiz-Muñoz G, Oguiza A, Recio C, Blanco J, Nimmerjahn F, Egido J, Gomez-Guerrero C: Fcγ receptor deficiency attenuates diabetic nephropathy. *J Am Soc Nephrol* 23: 1518–1527, 2012
 21. van Timmeren MM, van der Veen BS, Stegeman CA, Petersen AH, Hellmark T, Collin M, Heeringa P: IgG glycan hydrolysis attenuates ANCA-mediated glomerulonephritis. *J Am Soc Nephrol* 21: 1103–1114, 2010
 22. Novak J, Julian BA, Mestecky J, Renfrow MB: Glycosylation of IgA1 and pathogenesis of IgA nephropathy. *Semin Immunopathol* 34: 365–382, 2012
 23. Novak J, Tomana M, Matousovici K, Brown R, Hall S, Novak L, Julian BA, Wyatt RJ, Mestecky J: IgA1-containing immune complexes in IgA nephropathy differentially affect proliferation of mesangial cells. *Kidney Int* 67: 504–513, 2005
 24. Menni C, Keser T, Mangino M, Bell JT, Erte I, Akmačić I, Vučković F, Pučić Baković M, Gornik O, McCarthy MI, Zoldoš V, Spector TD, Lauc G, Valdes AM: Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. *PLoS ONE* 8: e82558, 2013
 25. Pivac N, Knezevic A, Gornik O, Pucic M, Igl W, Peeters H, Crepel A, Steyaert J, Novokmet M, Redzic I, Nikolac M, Hercigonja VN, Curkovic KD, Curkovic M, Nedic G, Muck-Seler D, Borovecki F, Rudan I, Lauc G: Human plasma glycome in attention-deficit hyperactivity disorder and autism spectrum disorders. *Mol Cell Proteomics* 10: M110 004200, 2011.
 26. Novokmet M, Lukić E, Vučković F, Đurić Ž, Keser T, Rajšl K, Remondini D, Castellani G, Gašparović H, Gornik O, Lauc G: Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* 4: 4347, 2014
 27. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, Takeuchi F, Nagano Y, Miyamoto T, Kobata A: Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* 316: 452–457, 1985
 28. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, Pučić-Baković M, Novokmet M, Mangino M, Thaqi K, Rudan P, Novokmet N, Sarac J, Missoni S, Kolčić I, Polašek O, Rudan I, Campbell H, Hayward C, Aulchenko Y, Valdes A, Wilson JF, Gornik O, Primorac D, Zoldoš V, Spector T, Lauc G: Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* 69: 779–789, 2014
 29. Mihai S, Nimmerjahn F: The role of Fc receptors and complement in autoimmunity. *Autoimmun Rev* 12: 657–660, 2013
 30. Malhotra R, Wormald MR, Rudd PM, Fischer PB, Dwek RA, Sim RB: Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* 1: 237–243, 1995
 31. Cook HT: Complement and kidney disease. *Curr Opin Nephrol Hypertens* 22: 295–301, 2013
 32. Rombouts Y, Ewing E, van de Stadt LA, Selman MH, Trouw LA, Deelder AM, Huizinga TW, Wuhler M, van Schaardenburg D, Toes RE, Scherer HU: Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. *Ann Rheum Dis* 74: 234–241, 2015
 33. Ercan A, Cui J, Chatterton DE, Deane KD, Hazen MM, Brintnell W, O'Donnell CI, Derber LA, Weinblatt ME, Shadick NA, Bell DA, Cairns E, Solomon DH, Holers VM, Rudd PM, Lee DM: Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum* 62: 2239–2248, 2010
 34. Tomana M, Schrohenloher RE, Koopman WJ, Alarcón GS, Paul WA: Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases. *Arthritis Rheum* 31: 333–338, 1988
 35. Tomana M, Schrohenloher RE, Reveille JD, Arnett FC, Koopman WJ: Abnormal galactosylation of serum IgG in patients with systemic lupus erythematosus and members of families with high frequency of autoimmune diseases. *Rheumatol Int* 12: 191–194, 1992
 36. Anthony RM, Ravetch JV: A novel role for the IgG Fc glycan: the anti-inflammatory activity of sialylated IgG Fcs. *J Clin Immunol* 30[Suppl 1]: S9–S14, 2010
 37. Kaneko Y, Nimmerjahn F, Ravetch JV: Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* 313: 670–673, 2006
 38. Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC, Ravetch JV: Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* 320: 373–376, 2008
 39. Huffman JE, Pučić-Baković M, Klarić L, Hennig R, Selman MH, Vučković F, Novokmet M, Krištić J, Borowiak M, Muth T, Polašek O, Razdorov G,

- Gornik O, Plomp R, Theodoratou E, Wright AF, Rudan I, Hayward C, Campbell H, Deelder AM, Reichl U, Aulchenko YS, Rapp E, Wuhler M, Lauc G: Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* 13: 1598–1610, 2014
40. Karsten CM, Pandey MK, Figge J, Kilchenstein R, Taylor PR, Rosas M, McDonald JU, Orr SJ, Berger M, Petzold D, Blanchard V, Winkler A, Hess C, Reid DM, Majoul IV, Strait RT, Harris NL, Köhl G, Wex E, Ludwig R, Zillikens D, Nimmerjahn F, Finkelman FD, Brown GD, Ehlers M, Köhl J: Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcγRIIB and dectin-1. *Nat Med* 18: 1401–1406, 2012
41. Ferrara C, Stuart F, Sondermann P, Brünker P, Umaña P: The carbohydrate at FcγmaRIIIa Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. *J Biol Chem* 281: 5032–5036, 2006
42. Pucic M, Knezevic A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redzic I, Campbell H, Wright A, Hastie ND, Wilson JF, Rudan I, Wuhler M, Rudd PM, Josic D, Lauc G: High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* 10: M111 010090, 2011.
43. Masuda K, Kubota T, Kaneko E, Iida S, Wakitani M, Kobayashi-Natsume Y, Kubota A, Shitara K, Nakamura K: Enhanced binding affinity for FcγmaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol Immunol* 44: 3122–3131, 2007
44. Moayyeri A, Hammond CJ, Valdes AM, Spector TD: Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol* 42: 76–85, 2013
45. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
46. Stevens PE, Levin A; Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members: Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med* 158: 825–830, 2013

This article contains supplemental material online at <http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2015010109/-/DCSupplemental>.

G.4 Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model

SCIENTIFIC REPORTS

OPEN

Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model

Received: 08 July 2016
Accepted: 28 October 2016
Published: 25 November 2016

Jonas Zierer^{1,2}, Tess Pallister¹, Pei-Chien Tsai¹, Jan Krumsiek^{3,4}, Jordana T. Bell¹, Gordan Lauc^{1,5}, Tim D Spector¹, Cristina Menni¹ & Gabi Kastenmüller^{1,2,4}

Although association studies have unveiled numerous correlations of biochemical markers with age and age-related diseases, we still lack an understanding of their mutual dependencies. To find molecular pathways that underlie age-related diseases as well as their comorbidities, we integrated aging markers from four different high-throughput omics datasets, namely epigenomics, transcriptomics, glycomics and metabolomics, with a comprehensive set of disease phenotypes from 510 participants of the TwinsUK cohort. We used graphical random forests to assess conditional dependencies between omics markers and phenotypes while eliminating mediated associations. Applying this novel approach for multi-omics data integration yields a model consisting of seven modules that represent distinct aspects of aging. These modules are connected by hubs that potentially trigger comorbidities of age-related diseases. As an example, we identified urate as one of these key players mediating the comorbidity of renal disease with body composition and obesity. Body composition variables are in turn associated with inflammatory IgG markers, mediated by the expression of the hormone oxytocin. Thus, oxytocin potentially contributes to the development of chronic low-grade inflammation, which often accompanies obesity. Our multi-omics graphical model demonstrates the interconnectivity of age-related diseases and highlights molecular markers of the aging process that might drive disease comorbidities.

Aging is a multi-factorial process that affects the entire organism, thus causing decreased fitness, disease and eventually death. As the population of western countries is aging¹, the prevalence of a variety of age-related diseases, such as cardiovascular disease, cancer² and chronic kidney disease (CKD)³ and many related diseases are increasing. Finding mechanisms that cause diseases with progressing age as well as better understanding disease comorbidity patterns is thus essential to counteract an explosion of health care costs. Epidemiological studies have already identified a broad spectrum of molecules associated with aging from various layers of biology accessible through modern omics technologies⁴. These molecules include epigenetic markers⁵, RNA abundances⁶, protein abundances⁷, post-translational protein modifications – such as protein glycosylation⁸ – and metabolite concentrations⁹. However, these studies analyzed omics datasets independently, thus neglecting the intrinsic interactions of biological entities within and across omics layers. Taking into account this complex interplay is necessary to unveil the causal structure of multi-factorial processes such as aging.

Various concepts have been proposed to integrate data from different molecular layers and (omics) technologies in systems biology and the newly emerging fields of systems genetics¹⁰ and systems medicine¹¹. Thereby, networks have been shown to be particularly useful to assess complex interactions in a dataset and to illustrate multivariate dependencies¹². As an example, using network approaches, it was demonstrated that co-occurring diseases are linked to mutations in the same gene¹³, in genes that interact with each other¹⁴, or in genes that are involved in the same metabolic pathway¹⁵, explaining observed patterns of comorbidity¹⁶.

¹Department of Twin Research and Genetic Epidemiology, Kings College London, London, United Kingdom.

²Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany. ³Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany. ⁴German Center for Diabetes Research (DZD), Neuherberg, Germany. ⁵Genos Glycoscience Research Laboratory, Zagreb, Croatia. Correspondence and requests for materials should be addressed to G.K. (email: g.kastenmueller@helmholtz-muenchen.de)

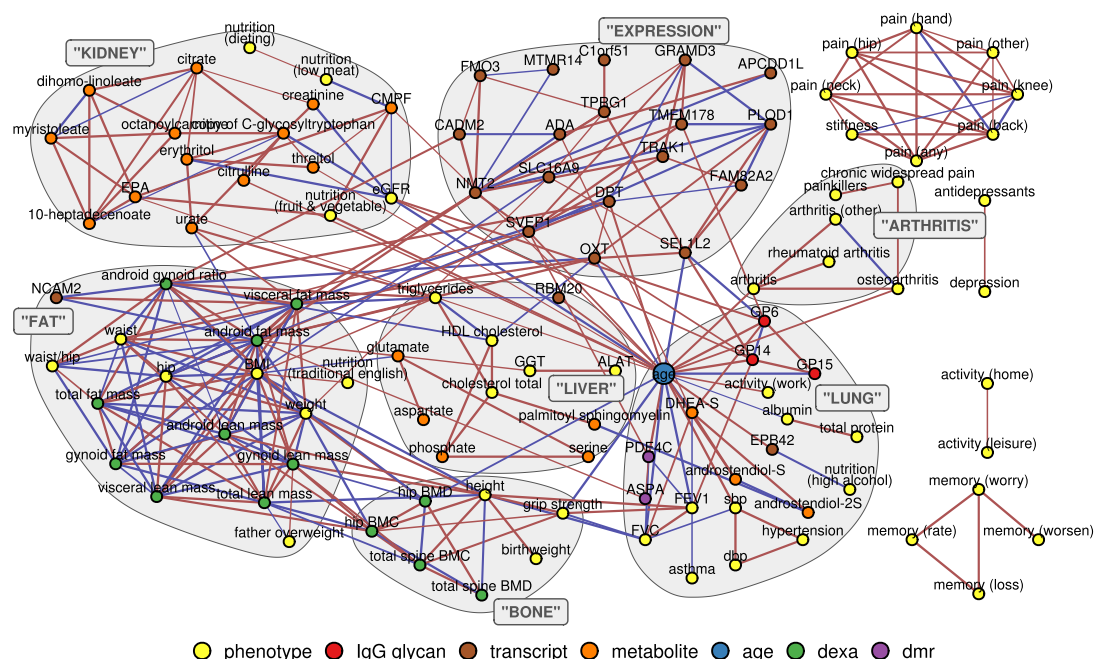


Figure 1. Multi-Omics MGM of Age. Each node in the graph represents one age-related variable. Omics markers were selected according to literature from epigenomics (purple), transcriptomics (brown), glycomics (red) and metabolomics (orange) datasets and combined with DXA measurements (green) and other clinical phenotypes (yellow). Edges between nodes were inferred using a mixed graphical model approach, and thus indicate the conditional dependence between variables; the color represents positive (red) and negative (blue) correlation. An unbiased cluster detection algorithm was used to identify densely connected modules within the network, indicated by grey borders.

Due to the increasing availability of high-dimensional omics datasets, networks can now be inferred also directly from measured data facilitating the unbiased analysis of specific conditions of interest, independently of prior knowledge. For instance, gene co-expression networks were used to analyze the influence of anti-cancer drugs on gene expression¹⁷. Integration of such co-expression networks with other omics layers allowed for prioritization of interesting, potentially causal, targets¹⁸. Even though these types of correlation-based networks led to a wide range of discoveries, they suffer from vast numbers of spurious correlations that inflate the number of edges and obscure the underlying mechanisms. Conditional independence graphs, such as graphical models¹⁹, were proposed as solution to overcome the problem of mediated associations²⁰ by revealing the relevant direct associations between variables. Although the direction of the associations and, thus, causality cannot be determined by these models in most cases, the resulting network of direct associations between variables can be considered as the undirected skeleton of their underlying causal structure. While the proposed graphical models are well established for multi-variate Gaussian distributed data, the extension to mixed distributions, as commonly observed in phenomics data (e.g. gender, disease states), is substantially more complex²¹.

In this study, we aimed to investigate the molecular basis of age-related diseases and its influence on disease comorbidities. To this end, we used an integrated mixed graphical model (MGM) approach to combine aging markers from four different high-throughput omics datasets on the same individuals, namely epigenomics, transcriptomics, glycomics and metabolomics, together with extensive phenotypic data. While we cannot infer actual causality using MGM, the resulting network of direct associations that are independent of all other variables within the model is expected to provide valuable insights into the direct molecular interdependencies between various age-related phenotypes. To the best of our knowledge, this is the first study that uses graphical models to combine data from multiple molecular omics and phenomics datasets.

Results

We inferred a mixed graphical model using observational data from a cohort of 510 women, aged between 34 and 84, integrating selected age-associated markers from four different omics datasets (see Materials and Methods) with 92 clinically assessed phenotypes (Supplementary Fig. S1). The final model consists of 145 nodes and 316 undirected edges connecting them (Fig. 1). Thus, it is much sparser (316 edges instead of 1900) than a regular correlation graph based on significant pairwise correlations of variables from the same dataset (Supplementary Fig. S2). Most of the nodes (96) form one large connected component, which we refer to as *age-mgm* in the following. There are two smaller components of 8 and 4 nodes that contain variables related to pain and memory function, respectively, two isolated pairs of nodes and 33 unconnected nodes. The degree, betweenness and clustering coefficients of all nodes in the network are presented in Supplementary Tables S1 and S2.

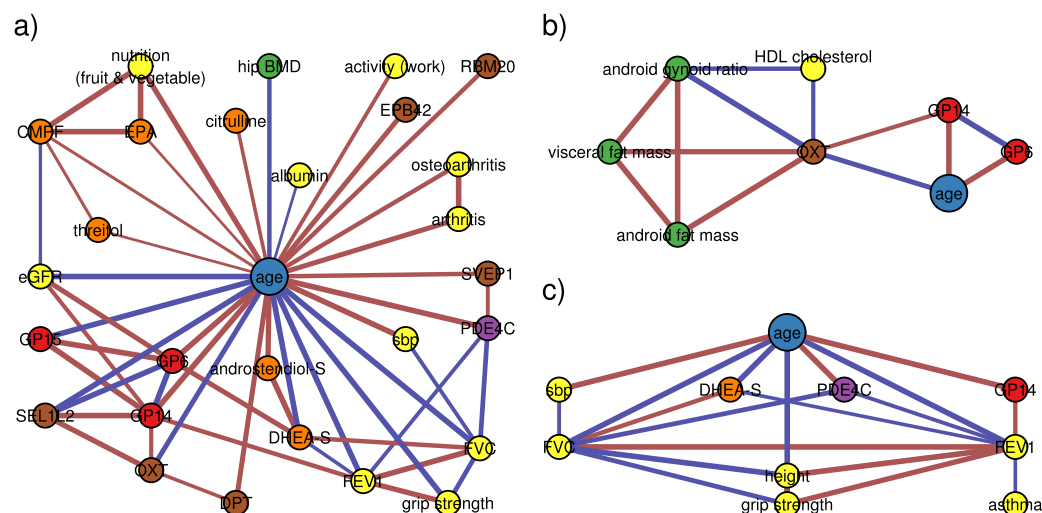


Figure 2. Selected modules from the graphical model. Each panel shows one subgraph from the age-mgm (Fig. 1). **(a)** The direct neighborhood of chronological age. **(b)** The hormone oxytocin (OXT) mediates association of fat mass variables with age as well as the IgG glycosylation marker GP14. **(c)** The direct neighborhood of the lung function measures *forced expiratory volume in one second* (FEV1) and *forced vital capacity* (FVC) contains three omics markers: dehydroepiandrosterone-sulfate (DHEA-S), phosphodiesterase 4 C (PDE4C) and the glycan peak 14 (GP14).

Topological Properties of age-mgm. The large connected component age-mgm contains 96 variables including age, along with variables from all four omics datasets, and 286 edges connecting them. It has an average node degree of 6.0, an average local clustering coefficient of 46.6% and an average shortest path length of 3.2. Also, its small world index, as defined by Humphries and Gurney²², is 6.1 and so the age-mgm can be considered a small-world network with high local clustering and short path lengths. Removing age from the network does not reduce the small worldness of the network. In comparison, the correlation graph, restricted to the same vertices as in the age-mgm, has a just slightly higher clustering coefficient of 57.0% despite the much higher average node degree of 31.2, which results in a small world-index of only 1.7.

As expected, age is the most densely connected node with a degree of 27 (Fig. 2a). It has a low clustering coefficient (8.0%) but high betweenness (47.5%) centrality. This indicates that age connects different clusters, while its neighbors tend to be unconnected. With an average shortest path length of 2.1 age is also the most central node in the age-mgm.

Modularity of the age-mgm. There are more edges between variables originating from the same omics dataset than edges connecting them. Particularly transcriptomics and metabolomics variables form dense clusters with 37 and 34 edges within them, respectively. In contrast, only 7 edges connect transcriptomics and metabolomics variables with variables from other omics sets. Similarly, the body composition variables measured by *dual-energy X-ray absorptiometry* (DXA) are densely connected with 45 edges between them (Supplementary Fig. S3).

In order to analyze the graph structure in an unbiased way we used a modularity-based algorithm for cluster detection. This approach yielded seven modules (Fig. 1). The first cluster (EXPRESSION) contains all but three gene expression markers. It is connected with neighboring clusters mainly via expression levels of OXT, which has 6 edges outside of its cluster (Fig. 2b), and SVEP1, which has the highest betweenness centrality (10.5%) within the cluster. The second cluster (LUNG) contains age and several of its direct neighbors from different omics layers. The lung function parameters *forced expiratory volume in one second* (FEV1) and *forced vital capacity* (FVC) are the most densely connected phenotypes in the cluster (degree 8 and 7 respectively). Both are embedded in a tight cluster with local clustering coefficients of 35.7% and 47.6%, respectively (Fig. 2c). Age is also connected to another small cluster of arthritis-related variables (ARTHRITIS). The body composition variables fall in two different clusters, one of them containing bone density-related variables (BONE) and the other fat and lean mass-related variables (FAT). While the BONE cluster is densely connected with the LUNG cluster, all connections between the FAT cluster and the LUNG cluster, containing the age variable, are mediated, mainly via gene expression variables from adipose tissue. The next cluster (LIVER) contains the liver markers *alanine-aminotransferase* (ALAT) and *gamma-glutamyl transpeptidase* (GGT) along with cholesterol and triglyceride levels and several amino acids. It also contains the gene expression marker of the RBM20 gene that mediates the connection of the cluster with age and the LUNG cluster. The last cluster (KIDNEY) contains mainly metabolite levels, but also markers for nutrition and a measure of renal function, the *estimated glomerular filtration rate* (eGFR). With 9 edges, C-glycosyltryptophan is central within the metabolite cluster. However, the eGFR (degree 7) is the main connection of the metabolomics cluster with age as well as IgG glycosylation markers. The only connections of the renal cluster with other clusters, apart from the LUNG cluster via age, are edges between urate and the FAT and LIVER clusters.

Robustness of the age-mgm. For estimating the robustness of our model, we inferred additional networks based on different cutoffs for edge inclusion and on changed selection of omics variables. Comparing these networks with our original model, we found node centrality as well as module assignments to be stable when varying the cutoff for edge inclusion (Supplementary Fig. S4). The modules identified in the original age-mgm remained stable even when including all available metabolomics variables with known chemical identity, i.e. when basing the inference procedure on 196 metabolites in addition to the 23 pre-selected ones (Supplementary Table S4).

To investigate the reproducibility of the inferred age-mgm for a different set of samples we determined two separate models from disjoint datasets incorporating the first and second twin of each family, respectively, and compared these models to the original network. The two resulting models reproduce 93.5% of all edges, with only 21 edges being unique to the initial model (Supplementary Fig. S5A, Supplementary Dataset S1). Moreover, in the models of the two disjoint datasets these unique edges just missed the edge inclusion cutoff of 80% in most cases, which is most likely due to the reduced power in the smaller datasets (Supplementary Fig. S5B).

Discussion

In this study, we inferred a robust graphical multi-omics model of age-related diseases by integrating disease phenotypes with molecular markers from four omics layers based on data available for 510 women from the TwinsUK cohort. Despite the sparsity of our model, which omits mediated associations, most variables form one connected component (age-mgm) consisting of seven modules. Interestingly, each of these modules represents a different aspect of aging, such as metabolic aging linked to decline of renal function (KIDNEY cluster). Other aspects of aging include the change in body composition, which can be divided in the change of fat and lean tissue (FAT cluster), along with the closely related changes of gene expression in adipose tissue (EXPRESSION cluster), on the one hand, and the decrease of *bone mineral density* (BMD) and *bone mineral content* (BMC) on the other hand (BONE cluster) (Fig. 1).

Our model illustrates multivariate dependencies of age-related diseases that potentially explain comorbidity patterns. Edges in our model represent conditional dependence between two variables, while the absence of an edge implies their conditional independence given all other variables in the model. Specifically, this means that previously observed age-associations of the variables, which are not directly linked to age in our model, occur due to the mediation by other variables between them. This differentiation between mediated and direct associations allows us to draw conclusions on underlying mechanisms even though the causal directions cannot be inferred. In the following section we will discuss some key findings from our aging model in detail. Figure 3 summarizes additional hypotheses derived from the model, which we will not further discuss for the sake of brevity.

Lung Function is a Central Aging Process. Lung function appears to be a central aging phenotype in our age-mgm. Both lung function measures, FEV1 and FVC, are directly connected with age and are, besides age, the most densely connected nodes in the LUNG cluster, connected with three different omics markers (Fig. 2c): (i) The metabolite dehydroepiandrosterone sulfate (DHEA-S) is one of the most abundant hormones in humans and well known to decrease with age²³ and was even suggested as an anti-aging drug²⁴. Moreover, DHEA-S has been found to prevent and even revert pulmonary hypertension in rats²⁵, suggesting a causal effect of DHEA-S on lung function. (ii) The methylation probe cg17861230 lies in the PDE4C gene, an enzyme that catalyzes the hydrolysis of cAMP. Expression levels of PDE4C were previously found to be associated with lung function²⁶. PDE4 is a potential target for drugs against COPD and a PDE4 inhibitor, Roflumilast, has been approved by the EMA for treatment of COPD²⁷. In this example, our graphical model approach indeed unveiled a known causal interaction of variables while removing less relevant mediated associations. (iii) Finally, the IgG glycosylation marker GP14 is connected to lung function in the age-mgm. GP14 is a glycan structure with terminal galactose, which is known to change the inflammatory state of IgG²⁸. While defects of general protein glycosylation²⁹ as well as an involvement of IgG³⁰ in COPD have been previously reported, glycosylation of IgG has so far not been associated with lung function. Our model suggests a contribution of IgG mediated inflammation and might help to unveil mechanisms of lung disease in dedicated experiments. As IgG glycosylation is also related with kidney function in our age-mgm as well as in previous studies³¹, this might provide an explanation for the comorbidity of lung disease and renal disease.

Decline of Renal Function Links Age with Metabolic Shift. The blood metabolome was shown to be strongly influenced by age in several studies⁹. In the age-mgm most of the age-associated metabolites (13) form one large cluster with only four of them being directly linked to age, while the remaining nine metabolites are only indirectly associated with age. For six of these nine metabolites the shortest path to age is through eGFR, a measure of renal function. Even though our model is undirected, age is the only non-modifiable variable in our model. We thus hypothesize that with increasing age renal function declines leading to the major shift in the aging blood metabolome, which possibly causes further diseases.

Urate Mediates Association of Renal Function with Body Composition. Urate mediates the connection of the KIDNEY cluster with FAT and LIVER clusters. Hyperuricemia has been previously reported to be associated with obesity, particularly increased visceral fat mass³², and increased triglyceride levels³³, which appears to be a direct association according to our model. Indeed, there is evidence that urate actually contributes to the development of obesity and diabetes, rather than being just a consequence of obesity: Elevated serum levels of urate were found to predict, amongst others, obesity³⁴ and diabetes³⁵. By knocking out the uric acid transporter SLC2A9 in mice, DeBosch and colleagues found that hyperuricemia causes several phenotypes of the metabolic syndrome, including obesity, dyslipidemia and hypertension³⁶. Administering a compensating treatment attenuated some but not all of the observed symptoms. Hyperuricemia is also a known comorbidity of renal disease, however the causal direction of this association is controversial³⁷. Renal disease and uremia were also shown to

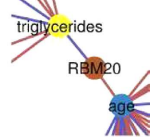
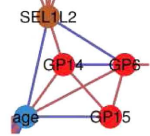
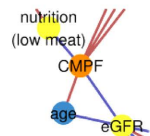
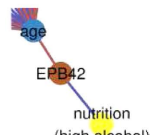
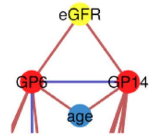
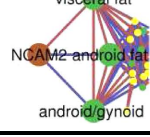
	Module	Description
<i>RBM20 mediates dyslipidemia with advancing age</i>		Expression of RBM20 mediates the association between triglyceride levels and age. Mutations of the RBM20 gene have been found to cause cardiomyopathy ⁷² . Our model indicates that this association might be mediated by dyslipidemia, which is a hallmark of cardiovascular diseases.
<i>SEL1L2 associated with IgG glycosylation</i>		SEL1L2 is directly connected to age and it is the only one directly linked to both, GP6 and GP14. Its function remains elusive, but its paralog, SEL1L, is essential for the degradation of misfolded proteins in the endoplasmic reticulum ⁷³ and particularly glycoproteins ⁷⁴ . This possibly causes the previously reported association between SEL1L and inflammatory bowel disease, and specifically Crohn's disease ⁷⁵ , which also is associated with glycosylation of IgG ⁷⁶ .
<i>CMPF mediates effect of low meat diet on renal function</i>		The metabolite 3-Carboxy-4-methyl-5-propyl-2-furanpropionic acid (CMPF) mediates the association between low meat diet and renal function. Consistently with our findings, CMPF was reported to be enriched by fish intake ⁷⁷ . It was also shown to be increased in patients suffering from chronic kidney disease ⁷⁸ , possibly causing renal function decline ⁷⁹ . Thus, lower levels of CMPF might explain the beneficial effect of dietary protein restriction on the progress of CKD ⁸⁰ .
<i>Adaption of cell membrane to alcohol intake</i>		EPB42, the only transcriptomics marker in the age cluster, is a transmembrane protein that has been associated with spherocytosis and osmotic fragility ⁸¹ . It is directly connected to high alcohol intake, which is known to affect cell membranes and might cause adaption of cell membranes ⁸² upon long-term exposure. Altered expression of EPB42 is possibly one such adaption.
<i>IgG-mediated inflammation associated with renal function</i>		The eGFR is directly linked to the IgG glycosylation traits GP6 and GP14, which aligns with previously published results on the involvement of IgG glycosylation in renal disease ³¹ and post-translational modifications in general ⁸³ .
<i>NCAM2 affects body composition</i>		The NCAM2 is the only omics variable located in the FAT cluster. Our model suggests that previous associations of mutations within NCAM2 and waist circumference ⁸⁴ are mediated by fat distribution. A recent study found that the knockout of a homolog gene in mice leads to a reduction of food intake and body weight ⁸⁵ , which is possibly related to its effect on olfaction ⁸⁶ .

Figure 3. Additional implications of the age-mgm. The figure summarizes selected conclusions drawn from our age-mgm that are not discussed in detail in the main text. Each panel shows a small excerpt from the network, restricted to relevant nodes and edges. Coloring is consistent with Fig. 1.

affect the gut microbiome composition³⁸, which is known to be strongly associated with obesity and other symptoms of the metabolic syndrome³⁹. Thus, the microbiome is possibly a hidden mediating factor, not included in our model, of the association between hyperuricemia and obesity. Even though its mode of action remains elusive, urate appears to be a key factor for the comorbidity of renal disease and obesity.

Hormone Expression Directly Associates with Body Composition. It is commonly known that BMI as well as waist and hip circumferences and body fat mass change with age. Nonetheless, we found neither of them directly linked to age in our model. Instead, all associations between age and the fat cluster are mediated. One of the paths connecting the FAT cluster with age is channeled via urate and renal function (as discussed above). A second path leads via the EXPRESSION cluster and, particularly, the expression of oxytocin (OXT) (Fig. 2b), which accordingly mediates 6.0% of all shortest paths in the model. OXT is also directly linked to HDL cholesterol levels. While adipose tissue was traditionally considered as storage tissue, it receives increasing attention as endocrine organ⁴⁰ that amongst others produces OXT. OXT is a hormone with a broad spectrum of functions, ranging from reproductive functions and control of social behavior⁴¹ to energy metabolism⁴². One common explanation for the influence of OXT on obesity is its effect on food intake⁴³, but there is also a diet-independent effect of OXT on the lipid metabolism in adipose tissue⁴⁴. Thus, OXT was suggested as drug against obesity and type 2 diabetes development and has been successfully tested in a first pilot trial⁴⁵. Our results indicate that the age-related change of body composition can, amongst others, be attributed to alterations of gene expression in adipose tissue and particularly to a change in OXT expression, independently of food intake. OXT might also drive common comorbidities of obesity by causing dyslipidemia, which in turn increases the risk of – amongst others – cardiovascular diseases⁴⁶.

IgG Glycosylation as New Mechanism of Obesity-Associated Inflammation. Obesity is known to be associated with chronic low-grade inflammation and activation of immune function⁴⁷, which is thought to be an important mediator between obesity and common comorbidities, such as type 2 diabetes⁴⁸. In our model the expression of OXT mediates the association of android and visceral fat mass with inflammatory IgG glycosylation. The influence of oxytocin on IgG might be mediated by IL6, which was found to be less expressed due to

OXT *in vitro*⁴⁹ and thus causes decreased IgG production in B-cells⁵⁰. Our study confirms an effect of increased fat mass on IgG, mediated by OXT, *in vivo*. Moreover, it provides evidence that OXT also affects IgG glycosylation in addition to its expression, thus altering its inflammatory potential. We hypothesize that this is a new mechanism of obesity-induced inflammation, which appears to be independent from previously identified pathways that are mediated by leptin or adiponectin⁵¹. Both of them are co-expressed with OXT in our data (Pearson correlation $r = 0.2$, $p = 8.1 \times 10^{-9}$ and $r = -0.29$, $p = 6.1 \times 10^{-17}$ respectively), but not associated with any of the IgG glycosylation markers.

We also found IgG-mediated inflammation being directly linked with renal function (Fig. 3), suggesting altered inflammatory potential of IgG as possible mechanism causing comorbidities of renal disease, obesity and related phenotypes of the metabolic syndrome. This supports the theory of “inflammaging”, which proposes chronic low-grade inflammation as mechanism that drives disease onset during aging⁵².

Limitations and Future Directions. Due to the limited availability of large multi-omics datasets and comprehensive collections of clinical phenotypes, our study is restricted by the relatively small sample size of 510 individuals and, more importantly, we were not able to get access to comparable data from an independent cohort to replicate our results. Also, all of our participants are female. As a consequence, our model and the conclusions drawn from it might be only partly transferable to the entire population. However, more and larger multi-omics dataset will be available in near future, for instance from the UK Biobank or the US Precision Medicine Initiative, which will facilitate subsequent studies using our multi-omics integration approach. For the time being, we could only demonstrate the stability of our results by inferring separate models from two disjoint sets of our own dataset that include only one twin of each twin pair, respectively (Supplementary Fig. S5). The limited number of samples also made prior selection of variables indispensable. This selection can be expected to influence the topology and modularity of the final network model. However, in our study, doubling the number of omics variables by not pre-selecting metabolites from the metabolomics data did result in very similar topology and module assignments in the model (Supplementary Table S4). Also, upcoming larger datasets will allow to overcome this limitation by reducing the dimensionality of the data without relying on variable selection based on prior association analyses. While stability selection controls the family-wise error rate (FWER) of edges in the step of network inference, stability selection cannot quantify the total uncertainty in the model and its downstream analyses. However, analyzing the sensitivity of our approach against variations in the inferred network model (e.g. through different threshold for the selection of edges) demonstrated the stability of our results (Supplementary Fig S4, S5 and Table S4). Finally, our approach allows to detect mediation by variables included in the model and thereby enables differentiation between direct and indirect effects, it does, however, not allow to infer causality. Thus, based on our model, we only hypothesize about causal directions. Mendelian randomization might enable inference of causal direction using SNPs as instrumental variables. Much larger sample sizes are needed than available for this study, though. Ideally, potentially causal edges in our model should be further investigated in dedicated functional studies or randomized clinical trials to establish causality and infer causal direction.

Conclusion

This is, to our knowledge, the first study integrating data from four omics technologies and clinical phenotypes using an integrated statistical approach. Despite the relatively small sample size, our model confirms causal mechanisms of disease, which have been previously found using highly specific experiments and clinical trials, purely based on observational data from a generally healthy cohort. Moreover, we uncovered several new potential mechanisms that might contribute to disease comorbidities. We found, for instance, urate as key factor connecting body composition and renal function, as well as several phenotypes of the metabolic syndrome. Moreover, by integrating multiple omics datasets, we find the hormone oxytocin as a central mediator that connects inflammation and obesity and, thus, supports the theory of inflammaging.

Our study highlights the importance and the feasibility of data integration across omics layers including phenomics while considering multivariate dependencies. In the future this will help to focus on few, interesting associations, which can then be specifically tested in model organisms and clinical trials. Eventually this will speed up drug discovery by excluding irrelevant pathways and potential drug targets early in the development and thus limiting the set of potential targets and reducing costs of drug discovery.

Materials and Methods

Study Population. We analyzed data from the TwinsUK cohort, a national register of 11,000 adult twins recruited as volunteers without selecting for any particular disease or trait. For this study we selected 510 female participants (62 monozygotic twin pairs, 116 dizygotic twin pairs and 154 singletons) aged between 34 and 84 (mean 59.0 ± 9.4) with measurements for epigenomics, transcriptomics, glycomics and metabolomics available. The study has been approved by the local St. Thomas' Hospital Research Ethics Committee and was carried out in accordance with the approved guidelines. All study participants provided written informed consent.

Data Acquisition and Processing. The phenotypic data was collected using questionnaires and anthropometric measures during hospital visits. Additionally, four different high-throughput omics datasets were analyzed. With several hundred measured metabolites, thousands of RNA transcripts and particularly hundreds of thousands of CpG sites, network inference is not feasible. We used a knowledge-driven approach to reduce the number of variables from each dataset. To this end, we selected only variables which were previously reported to be strongly (and independently) associated with chronological age as described in the following (and listed in Supplementary Table S1).

Epigenomics. DNA methylation levels were measured in adipose tissue samples using Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA) as previously described⁵³. Data was corrected for technical variation using the beta mixture quantile dilation (BMIQ) method and corrected for batch effects and bisulfite conversion levels using linear mixed effect models. Weidner and colleagues⁵⁴ showed that only three aging related differentially methylated regions (aDMRs) are enough to predict the chronological age with high precision. Those three sites, namely cg02228185 (in ASPA), cg25809905 (in ITGA2B) and cg17861230 (PDE4C), were selected for further analyses.

Transcriptomics. RNA abundance was measured in abdominal fat samples using the Illumina Human HT-12 V3 Bead chip as part of the MuTHER project as previously described⁵⁵. The probe intensities were adjusted for batch effects using linear models prior to analysis. A previous study found 188 genes (199 probes) significantly associated with chronological⁵⁶ age. We performed stepwise regression to select expression probes independently associated with age. This procedure left 24 probes from 24 different genes (see Supplementary Table S1 for full list) for further analysis.

Glycomics. For this study IgG glycans were measured in a high-throughput manner as described by Pucic and colleagues⁵⁷. Briefly, IgG was first isolated from 90 µl plasma, the attached glycans were released, labelled with 2-aminobenzamide and analyzed by UPLC. The according chromatograms were divided in 24 glycan peaks (GP), corresponding to 24 glycan structures. The data has been described in detail before⁵⁸. Glycan peaks were global normalized, log transformed and corrected for batch effects using ComBat. It has been shown that a linear combination of only three IgG glycan structures - GP6, GP14 and GP15 - explains 58% percent of the variance in age⁸ and furthermore correlates with several aging associated phenotypes. These three structures were selected for our network analysis.

Metabolomics. An untargeted LC/MS and GC/MS platform was used to measure metabolite spectra from plasma and serum samples, respectively. Metabolites were subsequently identified by Metabolon Inc., Durham, USA, using their proprietary database⁹. Metabolite levels were scaled by the run-day median, imputed using the run-day minimum, inverse normalized and corrected for batch effects using linear mixed models with the batch as random intercept. About the half of all known circulating blood metabolites were reported to be associated with chronological age in several large population studies⁹. We selected 22 of these metabolites, which were shown to be independently associated with age and together explain 59% of the variance of chronological age⁹.

Clinical Phenotypes. A total of 92 phenotypes was combined with the previously described omics data (listed in Supplementary Table S2). Besides the chronological age, we included 13 body composition variables, measured by *dual-energy X-ray absorptiometry* (DXA), as previously described⁵⁹. In addition to DXA measurements we included common body composition measures, such as height, weight, waist and hip circumferences and body mass index (BMI). Lung function was assessed by measuring the *forced expiratory volume in one second* (FEV1) and the *forced vital capacity* (FVC) using standard spirometry⁶⁰. Biochemical measures of *gamma-glutamyltransferase* (GGT) and *alanine aminotransaminase* (ALAT) were used to determine liver function. We furthermore used the CKD-EPI equation⁶¹ to estimate the *glomerular filtration rate* (eGFR) from serum creatinine as measure of renal function. Moreover, we included data from various questionnaires, assessing disease states, such as arthritis, asthma and chronic pain. Additionally, questionnaires were used to collect lifestyle parameters. Amongst others, we included data about physical activity and nutrition. Food intake data was collected using an established food frequency questionnaire⁶². Item frequencies were merged into 54 food groups and transformed into orthogonal patterns using principal component analysis⁶³. We used the first five principal components, which correspond to five different dietary patterns (Supplementary Table S3), in our model. A complete list of phenotypes is shown in Supplementary Table S2.

Data Pre-Processing. We excluded samples with more than 20% missing values and subsequently excluded variables with more than 20% missing values. Remaining missing values were imputed using the mice package⁶⁴. All continuous variables were inverse normalized and categorical variables were dichotomized. To account for family relatedness, we included one variable indicating a unique identifier per family during network inference and removed the according node from the network prior to analysis.

Data Availability. The transcriptomics and epigenomics data are available at ArrayExpress (accession number E-TABM-1140 and E-MTAB-1866, respectively). All other TwinsUK omics data are publicly available upon request on the departmental website (<http://www.twinsuk.ac.uk/data-access/accessmanagement/>).

Network Inference. The mixed graphical model was inferred using the Graphical Random Forest (GRaFo) method²¹ with the complementary pairs stability selection (CPSS) modification⁶⁵. Briefly, for each variable all remaining variables were ranked according to their conditional dependence assessed by the random forest variable importance. Consequently, two ranks were calculated for each pair of variables x and y : one based on the variable importance of x for the prediction of y and the other based on the importance of y for the prediction of x . The maximum (i.e. worse) of these two ranks was used as rank of the pair and the best ranking pairs were added as edges of the graphical model. This procedure was repeated for 100 random subsets of the data, each containing the half of all samples, and their complementary set containing the other half of the samples. The resulting 200 graphical models were combined using CPSS⁶⁵ to control the family-wise error rate (FWER). Edges which were contained in more than 80% of all complementary pairs were included in the final model, thereby ensuring FWER < 0.05. As effect estimators from random forest, and partial effects in mixed models in general, are non-linear

and depend on other variables in the model, there is no estimator for the sign of an edge in our model. We, thus, inferred the signs from regression models, regressing each variable against all others, for visualization purposes.

Network Analysis. We analyzed the graphical model as undirected, unweighted network $G = (V, E)$, consisting of a set of vertices V and a set of edges E .

Several measures were calculated to assess the centrality of nodes in the network. The degree of a node v is defined as the number of edges that contain this node, thus assessing its direct associations. The clustering coefficient is the proportion of edges within the neighborhood of v that are present in the network. It measures the centrality of v within its local neighborhood. In contrast, the betweenness centrality considers indirect associations of v and assesses its importance for the network integrity. It is defined as the proportion of all shortest paths that contain v ¹². Real-world networks often consist of densely connected modules, so-called clusters or communities that represent functional units within the network⁶⁶. We used an unbiased way to identify clusters within our model, independently from the type of a variable. To this end, we used the algorithm of Brandes and colleagues⁶⁷. It optimizes the modularity score that increases with the number of intra-cluster edges and decreases with the number of inter-cluster edges. Despite the high local clustering, many biological networks are characterized by short average path lengths between nodes. These networks are referred to as small world networks. This concept was formalized by Humphries²², who introduced the small world index for networks, that assesses the small-world-ness of a network by comparing its clustering coefficient and average shortest path lengths with an Erdős-Rényi random graph.

Network Stability. To test the robustness of our model we investigated the dependence of the network topology on the inference process.

Firstly, we assessed robustness of node centrality as well as module assignments when varying the cutoff for edge inclusion. To this end, we defined different models by including edges that are contained in 20%, 40%, 60%, 80% and 100% of the subsamples, respectively, where 80% corresponds to the original model. Additionally, we analyzed a weighted network⁶⁸ including all edges that were observed in at least one subsample. As a measure of stability of node centrality, we determined the correlation of node degrees and clustering coefficients between the original model and the model in the networks for different edge cutoffs (Supplementary Fig. S4). To assess the stability of module assignments, we calculated the adjusted RAND index⁶⁹ as a measure of similarity between the seven network modules of the original age-mgm with modules identified from the networks that were inferred based on different edge cutoffs. The RAND index assesses the similarity of module assignments by counting the agreements between two different module assignments and adjusting it for the number of agreements that are expected by chance. An adjusted RAND index of 1.0 indicates identity between the module assignments of two networks while values around 0.0 indicate dissimilarity of the assigned modules. In addition, we compared the adjusted RAND indices of the networks for the different edge cutoffs with the background distribution of 1000 randomly sampled module assignments (Supplementary Fig. S4C).

Secondly, we investigated the stability of the network and, in particular, the module assignments depending on the pre-selection of omics variables prior to the model inference. To assess the influence of this selection step on our results we inferred a second model from the same dataset but this time including all metabolomics variables with known chemical identity, thus, completely dispensing variable selection for the metabolomics data. The resulting graph consist of 341 nodes (145 of them from the original model and 196 newly added) connected by 1152 edges. 707 of these edges are amongst the new metabolites, 174 connect one new metabolite with one original variable and 271 edges are amongst original variables, of which 253 are also in the original model. The 63 edges that are missing in the large network compared to the original model are, on average, contained in 58% of the subsamples of the large network, suggesting that they were excluded due to the limited power. We find the added metabolites predominantly peripheral to the age-mgm, with 160 of the 174 edges connecting new metabolites with original variables being amongst metabolites and the remaining 14 with either blood lipid measures or renal function. A graphml file of the large network can be found in the Supplementary Dataset S1. To compare module assignments for the large network with the assignments for the original age-mgm, we restricted the large network to the nodes of the age-mgm. Edges in this network represent conditional dependence, given all other variable in the age-mgm and additionally given the 196 added metabolites. We assigned modules using the spin-glass algorithm implemented in the igraph package (as calculating the optimal modularity is computationally expensive for large networks). Module assignments are compared using adjusted RAND index and comparison of detailed module membership (Supplementary Table S4).

The stability of the model depending on the underlying sample sets was assessed by comparing our initial model with models inferred from two disjoint datasets containing either the first or the second twin of each family, respectively. Singletons were distributed randomly across both datasets. The resulting models are provided in graphml format in Supplementary Dataset S1.

All data was analyzed using R (version 3.1.2) along with the randomForest (version 4.6), igraph⁷⁰ (version 1.0.1) and ggplot2⁷¹ packages. The final network model is available as graphml file in the Supplementary Dataset S1.

References

- Office for National Statistics, Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2015. <http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2015>, (Date of access: 23/09/2016) (2016).
- Driver, J. a, Djousse, L., Logroscino, G., Gaziano, J. M. & Kurth, T. Incidence of cardiovascular disease and cancer in advanced age: prospective cohort study. *BMJ* **337**, a2467–a2467 (2008).
- Coresh, J., Astor, B. C., Greene, T., Eknoyan, G. & Levey, A. S. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third national health and nutrition examination survey. *Am. J. Kidney Dis.* **41**, 1–12 (2003).

4. Zierer, J., Menni, C., Kastenmüller, G. & Spector, T. D. Integration of 'omics' data in aging research: from biomarkers to systems biology. *Aging Cell* **14**, 933–944 (2015).
5. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
6. Yang, J. *et al.* Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci. Rep.* **5**, 15145 (2015).
7. Menni, C. *et al.* Circulating Proteomic Signatures of Chronological Age. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 809–816 (2015).
8. Kristic, J. *et al.* Glycans Are a Novel Biomarker of Chronological and Biological Ages. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **69**, 779–789 (2014).
9. Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int. J. Epidemiol.* **42**, 1111–1119 (2013).
10. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15**, 34–48 (2014).
11. Gustafsson, M. *et al.* Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Med.* **6**, 82 (2014).
12. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
13. Goh, K. *et al.* The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685–8690 (2007).
14. Oti, M. Predicting disease genes using protein-protein interactions. *J. Med. Genet.* **43**, 691–698 (2006).
15. Lee, D.-S. *et al.* The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. USA* **105**, 9880–9885 (2008).
16. Park, J., Lee, D.-S., Christakis, N. A. & Barabási, A.-L. The impact of cellular networks on disease comorbidity. *Mol. Syst. Biol.* **5**, 1–7 (2009).
17. Butte, a J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. USA* **97**, 12182–12186 (2000).
18. Huan, T. *et al.* Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol. Syst. Biol.* **11**, 799–799 (2015).
19. Lauritzen, S. L. *Graphical Models*. (Oxford University Press, USA, 1996).
20. Krumsiek, J., Suhre, K., Illig, T., Adamski, J. & Theis, F. J. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* **5**, 21 (2011).
21. Fellinghauer, B., Bühlmann, P., Ryffel, M., von Rhein, M. & Reinhardt, J. D. Stable graphical model estimation with Random Forests for discrete, continuous, and mixed variables. *Comput. Stat. Data Anal.* **64**, 132–152 (2013).
22. Humphries, M. D. & Gurney, K. Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLoS One* **3**, e0002051 (2008).
23. Orentreich, N., Brind, J. L., Vogelmann, J. H., Andres, R. & Baldwin, H. Long-term longitudinal measurements of plasma dehydroepiandrosterone sulfate in normal men. *J. Clin. Endocrinol. Metab.* **75**, 1002–1004 (1992).
24. Baulieu, E. E. *et al.* Dehydroepiandrosterone (DHEA), DHEA sulfate, and aging: contribution of the DHEAge Study to a sociobiomedical issue. *Proc. Natl. Acad. Sci. USA* **97**, 4279–4284 (2000).
25. Bonnet, S. *et al.* Dehydroepiandrosterone (DHEA) prevents and reverses chronic hypoxic pulmonary hypertension. *Proc. Natl. Acad. Sci. USA* **100**, 9488–9493 (2003).
26. Tang, H.-F. *et al.* The role of PDE4 in pulmonary inflammation and goblet cell hyperplasia in allergic rats. *Biochim. Biophys. Acta* **1762**, 525–532 (2006).
27. Calverley, P. M. *et al.* Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. *Lancet* **374**, 685–694 (2009).
28. Karsten, C. M. *et al.* Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of Fc γ RIIB and dectin-1. *Nat. Med.* **18**, 1401–1406 (2012).
29. U. Nihlén, P. Montn  mery & L. H. L. Increased serum levels of carbohydrate-deficient transferrin in patients with chronic obstructive pulmonary disease. *Scand. J. Clin. Lab. Investig.* **61**, 341–347 (2001).
30. O'Keeffe, S. *et al.* Immunoglobulin G subclasses and spirometry in patients with chronic obstructive pulmonary disease. *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.* **4**, 932–936 (1991).
31. Barrios, C. *et al.* Glycosylation Profile of IgG in Moderate Kidney Dysfunction. *J. Am. Soc. Nephrol.* **27**, 933–941 (2016).
32. Takahashi, S. *et al.* Close correlation between visceral fat accumulation and uric acid metabolism in healthy men. *Metabolism*. **46**, 1162–1165 (1997).
33. Giacomello, A., N. D. S. & CP, Q. Relation between serum triglyceride level, serum urate concentration, and fractional urate excretion. *Metabolism* **46**, 1085–1089 (1997).
34. Masuo, K., Kawaguchi, H., Mikami, H., Ogihara, T. & Tuck, M. L. Serum Uric Acid and Plasma Norepinephrine Concentrations Predict Subsequent Weight Gain and Blood Pressure Elevation. *Hypertension* **42**, 474–480 (2003).
35. Nakanishi, N., Suzuki, K. & Tatara, K. Alcohol consumption and risk for development of impaired fasting glucose or type 2 diabetes in middle-aged Japanese men. *Diabetes Care* **26**, 48–54 (2003).
36. DeBosch, B. J., Kluth, O., Fujiwara, H., Sch  rmann, A. & Moley, K. Early-onset metabolic syndrome in mice lacking the intestinal uric acid transporter SLC2A9. *Nat. Commun.* **5**, 4642 (2014).
37. Johnson, R. J. *et al.* Uric acid and chronic kidney disease: which is chasing which? *Nephrol. Dial. Transplant* **28**, 2221–2228 (2013).
38. Vaziri, N. D. *et al.* Chronic kidney disease alters intestinal microbial flora. *Kidney Int.* **83**, 308–315 (2012).
39. Parekh, P. J., Balart, L. a. & Johnson, D. a. The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease. *Clin. Transl. Gastroenterol.* **6**, e91 (2015).
40. Hauner, H. The new concept of adipose tissue function. *Physiol. Behav.* **83**, 653–658 (2004).
41. Bartz, J. A., Zaki, J., Bolger, N. & Ochsner, K. N. Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* **15**, 301–309 (2011).
42. Chaves, V. E., Tilelli, C. Q., Brito, N. A. & Brito, M. N. Role of oxytocin in energy metabolism. *Peptides* **45**, 9–14 (2013).
43. Lawson, E. a. *et al.* Oxytocin reduces caloric intake in men. *Obesity* **23**, 950–956 (2015).
44. Deblon, N. *et al.* Mechanisms of the anti-obesity effects of oxytocin in diet-induced obese rats. *PLoS One* **6**, e25565 (2011).
45. Zhang, H. *et al.* Treatment of Obesity and Diabetes Using Oxytocin or Analogs in Patients and Mouse Models. *PLoS One* **8**, e61477 (2013).
46. Arca, M. *et al.* Usefulness of Atherogenic Dyslipidemia for Predicting Cardiovascular Risk in Patients With Angiographically Defined Coronary Artery Disease. *Am. J. Cardiol.* **100**, 1511–1516 (2007).
47. Fogarty, A. W. *et al.* A prospective study of weight change and systemic inflammation over 9 y. *Am. J. Clin. Nutr.* **87**, 30–35 (2008).
48. Esser, N., Legrand-Poels, S., Piette, J., Scheen, A. J. & Paquot, N. Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes. *Diabetes Res. Clin. Pract.* **105**, 141–150 (2014).
49. Szeto, A. *et al.* Oxytocin attenuates NADPH-dependent superoxide activity and IL-6 secretion in macrophages and vascular cells. *Am. J. Physiol. Endocrinol. Metab.* **295**, E1495–E1501 (2008).
50. Maeda, K., Mehta, H., Drevets, D. a. & Coggeshall, K. M. IL-6 increases B-cell IgG production in a feed-forward proinflammatory mechanism to skew hematopoiesis and elevate myeloid production. *Blood* **115**, 4699–4706 (2010).

51. Tilg, H. & Moschen, A. R. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. *Nat. Rev. Immunol.* **6**, 772–783 (2006).
52. Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: A link with inflammaging. *Ageing Res. Rev.* **12**, 685–698 (2013).
53. Grundberg, E. *et al.* Global analysis of dna methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* **93**, 876–890 (2013).
54. Weidner, C. I. *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **15**, R24 (2014).
55. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
56. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14**, R75 (2013).
57. Pucic, M. *et al.* High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Mol. Cell. Proteomics* **10**, M111.010090–M111.010090 (2011).
58. Menni, C. *et al.* Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. *PLoS One* **8**, e82558 (2013).
59. Menni, C. *et al.* Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity* **24**, 1380–1388 (2016).
60. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–338 (2005).
61. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
62. Bingham, S. A. *et al.* Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr.* **4**, 847 (2001).
63. Teucher, B. *et al.* Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res. Hum. Genet.* **10**, 734–748 (2007).
64. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
65. Shah, R. D. & Samworth, R. J. Variable selection with error control: another look at stability selection. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **75**, 55–80 (2013).
66. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
67. Brandes, U. *et al.* On Modularity Clustering. *IEEE Trans. Knowl. Data Eng.* **20**, 172–188 (2008).
68. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
69. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).
70. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).
71. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2009).

Acknowledgements

We thank Kerrin Small for her advice on the analysis of transcriptomics data. The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an ERC Advanced Researcher CM is funded by the MRC AimHy (MR/M016560/1) project grant.

Author Contributions

G.K. conceived the study; G.K., C.M. and T.S. supervised the study. T.P., P.C.T., J.B. and G.L. collected and normalized data and J.Z. performed the analysis and prepared all figures. J.Z., G.K. and J.K. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: GL declares that he is a founder and owner of Genos Ltd., which offers commercial service of glycomic analysis and has several patents in this field.

How to cite this article: Zierer, J. *et al.* Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model. *Sci. Rep.* **6**, 37646; doi: 10.1038/srep37646 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016